

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 16:09:06 ; Search time 77 Seconds

(without alignments)
4102.031 Million cell updates/sec

Title: US-09-901-572a-4

Perfect score: 6413

Sequence: 1 MNISKKLKSYTLIGLAVFG.....APTKPAPKPAAPKPAPE 1224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organeller:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5389	84.0	1062	2	Q9ZNM1 mycoplasma
2	5385	84.0	1062	2	Q9ZNM1 mycoplasma
3	1733.5	27.0	377	2	Q9LCS5 mycoplasma
4	1643	25.6	376	2	Q9LCS4 mycoplasma
5	483	7.5	1029	2	Q9ZNM1 mycoplasma
6	481	7.5	964	2	Q9ZNM1 mycoplasma
7	337	5.3	1144	2	Q9ZNM1 mycoplasma
8	322	5.0	1629	2	Q9ZNM1 mycoplasma
9	315.5	4.9	1635	2	Q9ZNM1 mycoplasma
10	311.5	4.9	1635	2	Q9ZNM1 mycoplasma
11	259.5	4.0	1813	5	Q8ID65 mycoplasma
12	257	4.0	2752	5	Q8ICT8 mycoplasma
13	252	3.9	9439	16	Q8CPT76 mycoplasma
14	250.5	3.9	4727	2	Q8G892 mycoplasma
15	247.5	3.9	1029	2	Q52708 rickettsia
16	245.5	3.8	1487	16	Q8YK40 anabaena sp

17	244.5	3.8	880	16	Q8EM45 mycoplasma
18	242	3.8	3194	16	Q9ZLM3 mycoplasma
19	241	3.8	1698	17	Q8PMP7 mycoplasma
20	237	3.7	8591	5	Q8IB94 mycoplasma
21	235	3.7	1245	5	Q8IR23 mycoplasma
22	233.5	3.6	4199	16	P74440 mycoplasma
23	230.5	3.6	1265	16	Q8EY72 mycoplasma
24	230	3.6	833	2	Q9S594 mycoplasma
25	228.5	3.6	2737	16	Q8XP07 mycoplasma
26	226.5	3.5	1681	5	Q8IJ39 mycoplasma
27	225.5	3.5	3848	5	Q8IDR0 mycoplasma
28	225.5	3.5	4688	16	Q9P008 mycoplasma
29	225	3.5	3165	16	Q8RDO9 mycoplasma
30	223.5	3.5	1999	16	Q8YMB9 mycoplasma
31	223.5	3.5	2940	5	Q8IRP9 mycoplasma
32	223	3.5	2893	16	Q25063 mycoplasma
33	223	3.5	3078	5	Q8IRK4 mycoplasma
34	222.5	3.5	1166	16	Q8EUX7 mycoplasma
35	222	3.5	1297	16	Q8EY73 mycoplasma
36	222	3.5	2201	2	Q8GFA6 mycoplasma
37	222	3.5	3763	5	Q8T2A1 mycoplasma
38	221	3.4	1076	5	Q8IL74 mycoplasma
39	220.5	3.4	1565	5	Q8TLM2 mycoplasma
40	220.5	3.4	3207	5	Q8IRQ0 mycoplasma
41	220	3.4	1128	16	Q8EWE4 mycoplasma
42	219.5	3.4	1615	2	Q9KRA4 mycoplasma
43	219.5	3.4	2306	16	Q8P107 mycoplasma
44	219.5	3.4	2340	16	Q9ZD91 mycoplasma
45	219	3.4	1324	5	Q8IF08 mycoplasma

ALIGNMENTS

RESULT 1

ID Q9ZNM1 PRELIMINARY; PRT; 1062 AA.
AC Q9ZNM1;
DT 01-MAY-1999 (TREMURel. 10, Created)
DT 01-MAY-1999 (TREMURel. 10, Last sequence update)
DT 01-OCT-2002 (TREMURel. 22, Last annotation update)
DE 120kDa membrane protein.
GN M1.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=20278096; PubMed=10816462;
RA Yoshida S., Fujisawa A., Tsuzaki Y., Saitoh S.:
RT "Identification and expression of a Mycoplasma gallisepticum surface
RT antigen recognized by a monoclonal antibody capable of inhibiting both
RT growth and metabolism.";
RL Infect. Immun. 68:3186-3192(2000).
DR EMBL: AB023292; BAA74942.1; -;
SQ SEQUENCE 1062 AA; 115773 MW; 7B689B0A0AEF060 CRC64;

Query Match 84.0%; Score 5389; DB 2; Length 1062;
Best Local Similarity 86.1%; Pred. No. 7.5e-272;
Matches 1054; Conservative 1; Mismatches 7; Indels 162; Gaps 1;
QY 1 MNISKKLKSYTLIGLAVFGALGASGFGQSDKSNNTQLYNCAARTLDANSVLAIGQ 60
DB 1 MNISKKLKSYTLIGLAVFGALGASGFGQSDKSNNTQLYNCAARTLDANSVLAIGQ 60
QY 61 NGSLEFVNLVDVDFNTTANGTTRIKDSTFKPLYGDLSDDCGCVYKQIVSDYTSRN 120
DB 61 NGSLEFVNLVDVDFNTTANGTTRIKDSTFKPLYGDLSDDCGCVYKQIVSDYTSRN 120
QY 121 RFDOGRTRAYVALLVNDANVHLKRTITNSNIGNRNNSKEVIGVNDPAHVIRPTDGG 180
DB 121 RFDOGRTRAYVALLVNDANVHLKRTITNSNIGNRNNSKEVIGVNDPAHVIRPTDGG 180
144

```

OY 181 TKENFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
Db 145 ----- 144
OY 241 TNADGMPDCGNGQITNDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKT 300
Db 145 ----- 144
OY 301 SFOLDERINTNSRIGNRRNNNSKFYIGVDNPAHYIRFTDDGTFNFNTOTGGEIVNDFI 360
Db 145 ----- 144
OY 145 ----- 144
OY 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNAADGMPDCGNGQITNT 420
Db 199 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNAADGMPDCGNGQITNT 258
OY 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKTSFOLDERFVYPEWTGSEE 480
Db 259 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKTSFOLDERFVYPEWTGSEE 318
OY 481 NKNTIRLATGSLPSENERWILDPGTPQVTLKEDSVNVFSRLVNSLSFTGDSIYIF 540
Db 319 NKNTIRLATGSLPSENERWILDPGTPQVTLKEDSVNVFSRLVNSLSFTGDSIYIF 378
OY 541 GTSELPLSMYTSFPTRLSDTLALNOVKTTDIEASTDNGTTTGTADTSSGSGTGACT 600
Db 379 GTSELPLSMYTSFPTRLSDTLALNOVKTTDIEASTDNGTTTGTADTSSGSGTGACT 438
OY 601 GNTTNTSCFVSNPTLNTYRSFGIDSKPTSAKIDETWADPNVTEAIVAEYRGLGIONEL 660
Db 439 GNTTNTSCFVSNPTLNTYRSFGIDSKPTSAKIDETWADPNVTEAIVAEYRGLGIONEL 498
OY 661 PITNAGNPTIRNTIGCVGFTSGSRVYLKASVNGDORPTGNFPELVYFGYLYGQOTRTGT 720
Db 499 PITNAGNPTIRNTIGCVGFTSGSRVYLKASVNGDORPTGNFPELVYFGYLYGQOTRTGT 558
OY 721 FWGTYKLLNNSPYDVLADARVGETETNOFRSTSLTYPMWGGILIEBGRKRSNPTPIRAQ 780
Db 559 FWGTYKLLNNSPYDVLADARVGETETNOFRSTSLTYPMWGGILIEBGRKRSNPTPIRAQ 618
OY 781 GDPEBSRSIFQSGISDWTYIYIOSVGLFPGDGINNLNVGKASSFINSRNPNGLEMTAA 840
Db 619 GDPEBSRSIFQSGISDWTYIYIOSVGLFPGDGINNLNVGKASSFINSRNPNGLEMTAA 678
OY 841 TTYLRSGIAGLARTSGLENOQFETTHQVLSVSPGOFSIKNIRTIFFPGNOIWMYLFETNE 900
Db 679 TTYLRSGIAGLARTSGLENOQFETTHQVLSVSPGOFSIKNIRTIFFPGNOIWMYLFETNE 738
OY 901 NKSSTVYTLRLADSSNDPASSSPTSILIDVNEIGVILPLDLSFTYVAAGNVALLFSSN 960
Db 739 NKSSTVYTLRLADSSNDPASSSPTSILIDVNEIGVILPLDLSFTYVAAGNVALLFSSN 798
OY 961 PPSPSSTAYVNTFNONISDLAEBSGAKYTSDFMGTOFPEDEYLLONGFTSOVANEVY 1020
Db 799 PPSPSSTAYVNTFNONISDLAEBSGAKYTSDFMGTOFPEDEYLLONGFTSOVANEVY 858
OY 1021 NOSFTLSLVDFTPAAGTNRVYVDPDGNLTNOMLPLKVOIQYLDGKYIYAKLKNNNIYV 1080
Db 859 NOSFTLSLVDFTPAAGTNRVYVDPDGNLTNOMLPLKVOIQYLDGKYIYAKLKNNNIYV 918
OY 1081 FSYNNGALPSWVVPFAIGSTILGILAMITIIIGLAIIGVPLRORLQDQKGFTEFKKVDL 1140
Db 919 FSYNNGALPSWVVPFAIGSTILGILAMITIIIGLAIIGVPLRORLQDQKGFTEFKKVDL 978
OY 1141 TAAVGSVYKKIITQTNVKKKPPALGAGSGDKRPAAPAKPAAPAKSPASBPAPKPTG 1200
Db 979 TAAVGSVYKKIITQTNVKKKPPALGAGSGDKRPAAPAKPAAPAKSPASBPAPKPTA 1038
OY 1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224
Db 1039 PKSGAPTKPTAPKPAAPKPTAPKE 1062

```

RESULT 2

O9REM7 PRELIMINARY; PRT: 1062 AA.

AC O9REM7; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DI 01-OCT-2002 (Tremblrel. 22, Last annotation update)

DE Putative cytochrome related molecule A.

GN CRMA.

OS Mycoplasma gallisepticum.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2096;

RN [1]

RC STRAIN=R.

RX MEDLINE=99061189; PubMed=9846732;

RA Goh M.S., Gorton T.S., Forsyth M.H., Troy K.E., Geary S.J.;

RT "Molecular and biochemical analysis of a 105 kDa Mycoplasma

RU gallisepticum cytochrome (gapA).";

RL Microbiology 144:2971-2978(1998).

RN [2]

RC STRAIN=R.

RX PubMed=11083776;

RA Papadakis L., Troy K.E., Gorton T.S., Liao X., Geary S.J.;

RT "Analysis of cytochrome-deficient, GapA-negative Mycoplasma

RL Infect. Immun. 68:6643-6649(2000).

DR EMBL, AF214004; AAF25382.1; "

SQ SEQUENCE 1062 AA; 115803 MW; 7B6D0D9D683A095 CRC64;

Query Match 84.0%; Score 5385; DB 2; Length 1062;

Best Local Similarity 86.0%; Pred. No. 1, 2e-271;

Matches 1053; Conservative 1; Mismatches 8; Indels 162; Gaps 1;

```

OY 1 MNISKKLSYTLIGLAVFGLASAFSGFQOSKSDNTOVLNOARTIDANSVRLAGLGQ 60
Db 1 MNISKKLSYTLIGLAVFGLASAFSGFQOSKSDNTOVLNOARTIDANSVRLAGLGQ 60
OY 61 NGSLEFNTLRVDNFTTANGTITIKLDSFTKPLYGILSDDCGGYKVVQIYSDYTTSSN 120
Db 61 NGSLEFNTLRVDNFTTANGTITIKLDSFTKPLYGILSDDCGGYKVVQIYSDYTTSSN 120
OY 121 RFDQOTRAYVALLVNDENAVHLKRNINSTRIGNRRNNNSKFYIGVDNPAHYIRFTDDG 180
Db 121 RFDQOTRAYVALLVNDENAVHLKRNINSTRIGNRRNNNSKFYIGVDNPAHYIRFTDDG 180
OY 181 TKENFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
Db 181 TKENFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
OY 241 TNADGMPDCGNGQITNDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKT 300
Db 241 TNADGMPDCGNGQITNDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKT 300
OY 301 SFOLDERINTNSRIGNRRNNNSKFYIGVDNPAHYIRFTDDGTFNFNTOTGGEIVNDFI 360
Db 301 SFOLDERINTNSRIGNRRNNNSKFYIGVDNPAHYIRFTDDGTFNFNTOTGGEIVNDFI 360
OY 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNAADGMPDCGNGQITNT 420
Db 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNAADGMPDCGNGQITNT 420
OY 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKTSFOLDERFVYPEWTGSEE 480
Db 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHRIKTSFOLDERFVYPEWTGSEE 480
OY 481 NKNTIRLATGSLPSENERWILDPGTPQVTLKEDSVNVFSRLVNSLSFTGDSIYIF 540
Db 481 NKNTIRLATGSLPSENERWILDPGTPQVTLKEDSVNVFSRLVNSLSFTGDSIYIF 540
OY 541 GTSELPLSMYTSFPTRLSDTLALNOVKTTDIEASTDNGTTTGTADTSSGSGTGACT 600
Db 541 GTSELPLSMYTSFPTRLSDTLALNOVKTTDIEASTDNGTTTGTADTSSGSGTGACT 600

```

```

Db 379 GTSLESLWYSEPTRLDLDLALNQYKTDIDEASSFDNCTTNGTTTTTSSGSGAGT 438
QY 601 GNTNTNSQVSNNTLTMTYTSFSGIDSKPTSAKIDETNMADPNVTEARITAEYKLGIONET 660
Db 439 GNTNTNSQVSNNTLTMTYTSFSGIDSKPTSAKIDETNMADPNVTEARITAEYKLGIONET 498
QY 661 PITNAGNFIRNTIGVGFTSGSRVYLRASYNGDORPTGFOFLVFGYLGYOQPTGT 720
Db 499 PITNAGNFIRNTIGVGFTSGSRVYLRASYNGDORPTGFOFLVFGYLGYOQPTGT 558
QY 721 FWYGYTKLLNSPDYLDSPRVGTETNOFRTSLTYPVMGVLTREGARSNTPYIRAQ 780
Db 559 FWYGYTKLLNSPDYLDSPRVGTETNOFRTSLTYPVMGVLTREGARSNTPYIRAQ 618
QY 781 GDFPESRSIFOSGYSNTYETIOSVLCFPGDIRNNLVNGVASSFLNSNRNPGLEKIAA 840
Db 619 GDFPESRSIFOSGYSNTYETIOSVLCFPGDIRNNLVNGVASSFLNSNRNPGLEKIAA 678
QY 841 TTYLRSGIIGARTSGLPNOQFPGTHQVLSVSPGDOFSSIKNIRTIFFPGNQLVFLPTNE 900
Db 679 TTYLRSGIIGARTSGLPNOQFPGTHQVLSVSPGDOFSSIKNIRTIFFPGNQLVFLPTNE 738
QY 901 NKKSSVYTLRLADSSNPDASSSPSLIDVNEIGVILPLDINSFYTVAAGVNALEFSSN 960
Db 739 NKKSSVYTLRLADSSNPDASSSPSLIDVNEIGVILPLDINSFYTVAAGVNALEFSSN 798
QY 961 PGSPGYTAVNTFNONLSDIAFEGSGAKYNSDFMGTOIEKDEYLLONGTSQVARRFVT 1020
Db 799 PGSPGYTAVNTFNONLSDIAFEGSGAKYNSDFMGTOIEKDEYLLONGTSQVARRFVT 858
QY 1021 NOSFLNSLVDFTPANAGTNRVAVDPDGNLTNOLPLKVQIQYLDGKYYPDAKLNNNLYT 1080
Db 859 NOSFLNSLVDFTPANAGTNRVAVDPDGNLTNOLPLKVQIQYLDGKYYPDAKLNNNLYT 918
QY 1081 FSTNNGALPSSWVVPVPAIGSTLGLAIMITLGLAIGIPRAOKRLDQKGFYTKKVDLT 1140
Db 919 FSTNNGALPSSWVVPVPAIGSTLGLAIMITLGLAIGIPRAOKRLDQKGFYTKKVDLT 978
QY 1141 TAAVGSVYKKIITQTNANVKKKPPALGKSGDKKPAAPAAAPAPAPKSPAKSPAKPTG 1200
Db 979 TAAVGSVYKKIITQTNANVKKKPPALGKSGDKKPAAPAAAPAPAPKSPAKSPAKPTG 1038
QY 1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224
Db 1039 PKSGAPTKPTAPKPAAPKPTAPKE 1062

RESULT 3
QYLC55 PRELIMINARY; PRT; 377 AA.
AC QYLC55:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE 120-kDa membrane protein MG3 (Fragment).
GN MG3.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S6;
RA Yoshida S.;
RT "Mycoplasma gallisepticum S6-strain gene encoding a 120-kDa membrane protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB033310; BAA94277.1; -.
FT NON_TER 377
SQ SEQUENCE 377 AA; 41773 MW; 417861CDE2B17037 CRC64;

Query Match 27.0%; Score 1733.5; DB 2; Length 377;
Best Local Similarity 65.2%; Pred. No. 1.7e-82;

```

```

Matches 352; Conservative 9; Mismatches 16; Indels 163; Gaps 3;
QY 1 MNTSKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDOLVNOARTLDANSVRLAGLG 60
Db 1 MNTSKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDOLVNOARTLDANSVRLAGLG 60
QY 61 NGSLEFNTVLRDVDDNFITTAANGTILKIDSTKPLYGIDLSDDCGGYVKQIVSDYTTSRN 120
Db 61 NGSLEFNTVLRDVDDNFITTAANGTILKIDSTKPLYGIDLSDDCGGYVKQIVSDYTTSRN 120
QY 121 RFDQROTAVYALLVNDDEAVHLKRINTNSNRIGNRNNSKFVIGVDNPAHYIRFTDDG 180
Db 121 RFDQROTAVYALLVNDDEAVHLKRINTNSNRIGNRNNSKFVIGVDNPAHYIRFTDDG 180
QY 181 TKFNFTNTOGELVNDPILDAPILPKDHPDWNLVYORKILPNDVNTAVPMPVGVSG 240
Db 181 TKFNFTNTOGELVNDPILDAPILPKDHPDWNLVYORKILPNDVNTAVPMPVGVSG 240
QY 241 TNADGMFDCNGQITNTDPIAQTKTNTDQNSFTNSGAMPGANRRYDSOLVKKRIKT 300
Db 241 NSATDGTDFRGNCG-----MPSAN----- 258
QY 301 SFQDERINTNSNRIGNRNNSKFVIGVDNPAHYIRFTDDGTFKNFTNTOGELVNDPIL 360
Db 259 ----- 258
QY 361 LDAPILPKDHPDWNLVYORKILPNDVNTAVPMPVGVSGTNADGMFDCNGQITNT 420
Db 259 ----- 258
QY 421 DPIAQTKTNTDQNSFTNSGAMPGANRRYDSOLVKKRIKTSTFQDEKFEVPEMTGSEE 480
Db 259 DPLATSKSSQSN-NPSSFNSSAQGTNRNRDSENVNHRKIKTSQDEKFEVPEMTGSEE 317
QY 481 NKNITRLATGSLPNSNRYWLDIPGTPQVTLKEDSVNVEBRLVNSNSLFTIGDSIYIF 540
Db 318 NKNITRLATGSLPNSNRYWLDIPGTPQVTLKEDSVNVEBRLVNSNSLFTIGDSIYIF 377

RESULT 4
QYLC54 PRELIMINARY; PRT; 376 AA.
AC QYLC54:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE 120-kDa membrane protein MG3 (Fragment).
GN MG3.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-F;
RA Yoshida S.;
RT "Mycoplasma gallisepticum F-strain gene encoding a 120-kDa membrane protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB033211; BAA94278.1; -.
FT NON_TER 376
SQ SEQUENCE 376 AA; F99BF61049E0D068 CRC64;

Query Match 25.6%; Score 1643; DB 2; Length 376;
Best Local Similarity 62.2%; Pred. No. 8.6e-78;
Matches 336; Conservative 13; Mismatches 27; Indels 164; Gaps 3;
QY 1 MNTSKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDOLVNOARTLDANSVRLAGLG 60
Db 1 MNTSKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDOLVNOARTLDANSVRLAGLG 60
QY 61 NGSLEFNTVLRDVDDNFITTAANGTILKIDSTKPLYGIDLSDDCGGYVKQIVSDYTTSRN 120
Db 61 NGSLEFNTVLRDVDDNFITTAANGTILKIDSTKPLYGIDLSDDFAGYVKQIVSDYTTSRN 120

```


Db 975 PQMLDA-----NKKDGA-----SSPSKPSAPAAKPKAGPTKP-SAPGAKPTAPAKPKAPART 1025

RESULT 6

ID 086075

AC 086075 PRELIMINARY; PRT; 964 AA.

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Adherence protein.

OC Mycoplasma gallisepticum.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

NCBI_TaxID=2096;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN=56;

RA Goh M.S., Gorton T.S., Forsyth M.H., Troy K.E., Geary S.J.;

RT "Molecular and Biochemical Analysis of a 105 kDa Mycoplasma

gallisepticum cytohesin.";

RL Microbiology 0:0-0(1998).

DR EMBL; AF083976; AAC32609.1; -

SQ SEQUENCE 964 AA; 105556 MW; 8E20F1C09F45B634 CRC64;

Query Match 7.5%; Score 481; DB 2; Length 964;

Best Local Similarity 22.6%; Pred. No. 7.4e-17;

Matches 270; Conservative 140; Mismatches 386; Indels 398; Gaps 55;

QY 110 QIVSDYTSNRNRFQDQRT-----RAYALLVNDENYHLKRINTNSNRIGN 155

Db 82 ETVYDSDQNKIPKPKTKLLDSESESESGINGRTY--ANINQKNLQGVIVKYNELPNS 139

QY 156 RN-----NNSKEYIG--GVND-PAHYTRFTDDGTEKEN---FTNQTGCEIYNDFTLDAPILP 205

Db 140 ENPFAVENNAFIKPKIMVDNYPSTWQSGANGKMFVLFQFYKKNDPNAVNRRYRAKYP 199

QY 206 KDL-----HPDWNYLYIQRIKLPNDVNTAVVWP--VGRVSGTNADG 246

Db 200 KRLQIOTTPPLDSSSPYEHPEWY-----EGNOFVPMQOYITNLGLYAKDG 248

QY 247 MEDC--GNGOINTNPDIQTKTTTNDQNPSTFNSGAMPGANNRYDSQLNKHRIKTSFOL 304

Db 249 MYLFGNGCTWVW-----NNSALSIGV-----FRTRKE- 276

QY 305 DERINTNSNRIGNRNNNSKFVIGVUNPAHYTRFTDDGTEKENFTNQTGCEIYNDFTLDAP 364

Db 277 -----NRTAEPQNTKTIV--GYRYGILLSAISFDATF-----NGLALAP 314

QY 365 ILPKDLHPWNYLYIQRIKLPNDVNTAVVWP--VGRVSGTNADG 421

Db 315 ALGQ-----DVGHEVPRLAIVGVSSPRGANGNIFLGSALTWTGTINGG 356

QY 422 PLAQTK--TTNDQNPSTF-----NSGA-----MPGANNRYDSQLNKHRIK 461

Db 357 NPLDTKWHSPAVIEDAPTTFTITVNSSGVQLNNGSQGSTSPMNSNG----- 403

QY 462 TSPQDLDEKFFVPEPMTSSEENKNTTRLATGSLPSENERWIIDIGTPOVTLKEDSVNF-S 520

Db 404 -----NESIPY--RWINSYDYSVRFALISKPAG-----GNTKQVE-----SLFTT 443

QY 521 RLYLNSVNSL--SFIDSIYIEGTSPLMYWSPFTRLSDLLALNOVKTDLEASSTDN 578

Db 444 ALKLDLNLSPNKFTEQNNIFSYAML----- 470

QY 579 GTTNGCTTTADTSSGTAGTGNNTNTSQTAVSNPPLNTYRSPGIDSKPTSANKIDETNW 638

Db 471 -----DGRQMSLGRKDSWTLTNTINNFTYNTQQLASTAAGENAN----- 512

QY 639 ADPNVTEARLYAEVRLGIONEIPITNAGNFRITNTIGCVGFTSTGSKRYVLASTYNGDQRP 658

Db 513 -PRNIIINA-----LTTAKGFRDRDIGNVYTY-----S 539

QY 699 GNFQPLVYEGYLGYYQOTRTGTFW---YGYKLLNNSPYDVLDSPRVGTETNOFR-RTSL 754

Db 540 NNTNKTFTY-----YVGGAITTWPEVQVNYKTSANITTYNLRTDPGCTPATQDANTY 594

QY 755 TYPVMGYLTEREGARS--FSNTPYIRAGDDTPESRSIFOSGYSDNTYEXYIOSYLGPDGIR 812

Db 595 SSKLNGAVILYSTQDQGWYNGSIYVKASPTPSQ-----GYTWQOFKGLTTASNAVIS 649

QY 813 NNLNVKAKSSFLNSRPNPNCLEMTAATYILRSQIGLARTSLPNOQPPGTHQVIVS 872

Db 650 NMTKACY-----SIRPDD-----TVFSYSKI-----PF----- 673

QY 873 PGDOFSSIKNIRTFPGNOLWYELFTENNKKSSVYTLRLADSSNPDPASSFSPSTLIDV 932

Db 674 -EKEITAAVNRSLDS-----YYQLNGEISVNTVARY--SPSSA-----LTLL 715

QY 933 EIGVILPLDSEFYTVNAAGNVALFSSNPSPGSYTAVNTFNQNLSDIAFEESGARYTSD 992

Db 716 PKRITNPMLNRD-----NVIGQCAFISRN-DIPSSF-----FEKKINDIV----- 754

QY 993 FMCTIQPKDEYILQNGFTSOVARNFVITNSFLNSLVDETPA--NAGTNYRVVYDDGNT 1050

Db 755 -----TTADGTEVLDSKYINSITRYTPPONNPHIRLRLIDRSRA 796

QY 1051 TNQNLPLKYOIQYLDKYYDAKLKNNNLVT--FSYNNFGALPWWVPTAIGSTGLIAMI 1109

Db 797 TDFITLILQV-LVDEEYVAVPQANSVYSDQFTGFDALPGVLYVALSIPITIALAL 855

QY 1110 ILGLAIGIPLRAQRKLDQKFKTTFKKVDTLTAAGSVYKKIITQTA--NVKKRPAALGA 1167

Db 856 ALGLIGIPMSQNRKMLKQGFALISNKKVDILTAAVGSVFQOITNRTSVETIKKTPQLQA 915

QY 1168 GKSGDKKPAAPAAKPAAPARPAKASPAKPTPKGAPTRKPAK-KPAKPT 1220

Db 916 ---NKKDGA---SSPSKPSAPAAKPKGPTKP-SAPGAKPTAPAKPKAPART 960

QY 950371 PRELIMINARY; PRT; 1144 AA.

AC Q50371;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE Pl-like adhesin.

OC Mycoplasma plium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

NCBI_TaxID=2122;

RN [1]

RC SEQUENCE FROM N.A.

RX MEDLINE=94131957; PubMed=8300531;

RA Thiam T.N., Ferris S., Bahnaoui E., Canarelli S., Montagnier L.,

RA Blanchard A.;

RT "Molecular characterization of the Pl-like adhesin gene from

Mycoplasma plium.";

U. Bacteriol. 176:781-786(1994).

DR EMBL; L19685; AAC36866.1; -

SQ SEQUENCE 1144 AA; 126732 MW; E7801C55C102D930 CRC64;

Query Match 5.3%; Score 337; DB 2; Length 1144;

Best Local Similarity 21.5%; Pred. No. 2.8e-09;

Matches 299; Conservative 168; Mismatches 491; Indels 434; Gaps 74;

QY 5 KKLK---SYTLIGLAVFALGASAFGEKOSKSDNTQLVNO-----ARTLDANSVL 55

Db 2 KRIKFNKYLILSLVSTTVISAAISLSTFKNQDQISNPIINQVASFNSPSIVGNKVGK 61

QY 56 AGLGQNSLFTNYLROVDDNFITRANTTIKIDSFTRPLGLD-----LSPD-----CGGY 106

Db 62 IRHMÖNNN-FNGVEIKNGEFVLTSTQSATRIDAFGLIWEPEKIASDSOYANLAGK 120

QY 107 KVRQIVSDYTSNRNRFQDQRTRAYVALV-----NDEANVHLKRINTNSNRIGNRNNNSK 161

Db 121 KVEETLODEGDSN-----LXLLLPKNTPDOKASIDEPDLYAVNELTSGSSKQ 172

Qy 162 FVI-----GGVDNPAHVIR--FTDDGRKF--NEFNOTGEIVDFILDPILPK 206

Db 173 AFVVOIIEENVNLQOOSTWPSFTIKGLMHIDPKMNDYFNOMKSSQSS-----STFFLK 227

Qy 207 DLHPWYNLYIORKILPN-DVNT-----AVVWP-----VGRSGTNAADGMDC 250

Db 228 EDHPSTVYANNSEKXHNADQNTONGNKGSAVVLPWKQYITNLGNMKAQVILIEF- 286

Qy 251 GNGQITNDPLAQ-----KTTDNONPSTFNSGAMPGA-----NNRYSQNLV 294

Db 287 GNGSIYN-DPEALISGMKMLDKFAPYSGNDINNO-----NGGIFAYALLRKYDPS----- 338

Qy 295 KHRKTSFOLDERTINTSNRIGNRNNNSKEVYIGVDNPAHVIRFTDDGTGFENTNTOGE 354

Db 339 KPLIGTS-----APNR-----RWNOSAPIGQDNETYPRILAVGGVAINMST----- 382

Qy 355 IVNDFILDAPILPKDLHPWYNLYI-----QRKILPNQNTAVVWPVGRVS 401

Db 383-----EATYLYTLAAGITVGAKEASAREVINS--NT-----S 413

Qy 402 GTNADDMFGDCNGOITNTDPLAQTKTTDNONPSTFNSGAMPANNRYDSOLVVKRIK 461

Db 414 TNKVYTKIQOKRSJOLTLGANTITNFKDT-----AANSIDAL-----LFG 453

Qy 462 TSFOLDEKFPYFEMTSGSEENKNTRLATGSLPSNEREYILIDPGTPVYTLKEDSVNFSR 521

Db 454 TAFNIDSLINLP-----TKLENLITIQ--NVPOYESF--DVGATMSVSAAGTYFYFDK 506

Qy 522-----LYLSVNSLSFGDSIYIFGISELPSLMTYSEFPRLDLMALNOVTDIEASST 576

Db 507 KMHASSSTTIDNTYTTASNGWNNGRTAF--WSY-PN--NDIGSIQPKTND----- 555

Qy 577 DNGTTNGTITTTADTSSGTCAGTGNTNTSQTASNPTLNTY--RSEGDISKPTSA--NK 632

Db 556 NNNAAT-----YNNLSLJEN-AIYYXSTLSFEGYSLKCGJLTK 594

Qy 633 IDETNWADPNVTEARIVAEYRLGIONEIPITMAGNFIRNTIGGVFTSGSVYLRASYN 692

Db 595 IEMPSEKNP-----ENTITGYAM-QVGRKSIYVLINEPK 625

Qy 693 GDORPFGNPOPELYFEGYLGQOTRTGTFWYG-----TYKLANSPYVDLDSPRVGTETNQ 748

Db 626 SDRKSLAIHGPPSSISIG--ESNLYGSAXKGDMDYVYKINNSNIGYVPSD--YSNITNN 680

Qy 749 FRRTSLTLYPMGQYLTGEGARFSNT--PYTAAQDPTESNSIPQSGYSDN-----TYEYI 802

Db 681 IINTGVAL-----YVY--GIKDFNTIPTLASOE-----IGNSPREDSSUTTKNGTL 727

Qy 803 QSVYL--GPDGIRNNLTNVGVKASS--FLNSNRPNPNEGEMIAATYLRSGIQLARTSGLEN 858

Db 728 OPTIPWMDYFGL--NSTNPFNSELSLMLNNOCTKTNNE----- 764

Qy 859 QQPRGTHQVIVSPGDQFSSIKNTITIFPNQMLV-----YFLFTNENN-----KSSVYN 908

Db 765-----HFIVTKSP--EISE-----YGAALTEFPIYNGSSNNNAKKSGKRAMFE 808

Qy 909 LRLADSSNPQASSFSTSLIDVNEIGVILPLIDNSFTV-----N 949

Db 809 VK-----DSNLSMSTTVQMOYQGLDLSNLTADSYVOKUNNEPOGDPDVLLKTRDDT 860

Qy 950 AAGNVALEFS--SNPGSFG--SYTAVNTFNQNLSDIAFEGSGAAYTSDPMGTQFQDEYLI 1006

Db 861 QKNNDIFGQJNTRREGISYCKL--KON-----YGSYFETISEIDRLSL 905

Qy 1007 ONG--FTSOVAANFYTN-----QSFINSIVDFT--PA 1034

Db 906 GNGQFVNNLSNOLILNLIANLIYVNLSTYGNPLDSSKSTIRIVKNOLENEFYQVTKNDI 965

Qy 1035 NAGTYNR-----VVPDPCNLTNQLPLKVOIOTYLDGKYDADAKLKN-----NLVYFS--YNN 1085

[illegible]

```

Db      880 TGGG-----PRTLLDQANL-OLMTGAGWRNDRKSSGSGSDENHFTKTSATGMDQOG 928
QY      595 SNGAGNTNTNTSQ---TVSNPLTNTYRFBGIDSKRTSKANKIDETWMAD--PVIATARI 649
Db      929 QSGTSAGNDSDLSKODKISKSGDSLTTQ---DGNATGOQ--EATVYTNLPPMLTPT--- 978
QY      650 AEXRLQIONEIPTTNAGN-----FIRNTIGVGFTSGSRVYLRASYNDGDPPTNGFOP 703
Db      979 ADM---PALSTFTNNNNORAOLEFRLGLGSI-----PVLVNSSGSSNNKFOATDOK 1027
QY      704 FLYVFGYLYQOQTRIGTFWYGYK-LLN-----NSPYVDLDSPRYG---TE 745
Db      1028 WSYT--DLQSDQTKMLPAYGEVNGLLNPALETYFNTFRAGSGSNTTSSPGIGKRIPE 1085
QY      746 TNOFRRTSLTYP-----VMGGLTEBARSFNSPTYIRAO----- 780
Db      1086 QNDSKATLTTPCLANTPODVGNLVSGTSLFQJLGGWL------SFTDFIKPRAGYL 1138
QY      781 -----GDPESRSIFOS-----GYSDNTYEIOGVLGFDGI-RNNLNVGKASS 823
Db      1139 GJOLSGLDASDSQRELINAKRPMWAFRGSWVRRLGRVESVMDLKGVMDQADQALAAQAT 1198
QY      824 FLNSN--RPNPGL---EMIAATY--LRQJGLARTS----- 854
Db      1199 SEASGALAPHPMALAFQVSYVEASVSSSTSSGSGSSNTSPYLHLIKPKKEVSTQL 1258
QY      855 --GL-----PNO-----QPFCTTHOYISVSPGDQSSIKNITTTPGNOQLFTLTNEN 901
Db      1259 DOGLKMLDPNOVTRKLRQSFQDHS------QPOSILKTTTPVPG-----TS6G 1302
QY      902 NKSSVYTLRLADSSNPDASSFSPTSLIDYNEI---GVILPLLDSEFTVNAANVALF 957
Db      1303 NIGSVLS-----GGAGGGSSGSGCVLSPYERISGMLVGLPST-----SDGNTS-- 1350
QY      958 SSNPGSPGTYAVNTNQ-----NLSDIAEESGARYTSDFWGTIOFKPDEYLLIONGFTS 1012
Db      1351 STNNLAPNT---NTGNDVYGVGRLS---ESNAKMNDDVDGIVRPLAELLDEGEQFA 1402
QY      1013 QVARNVT-----NOSFLNSLYD-TPRAN-----AGTNRV-----Y 1043
Db      1403 DTGPOSVKRKSPOQIDFNRLTTPVTDLPDTMLVYDOYIPLFDIPASVNRKMWRLKY 1462
QY      1044 VDPDGNLTNQNLPLKYOIOLDKYDYDAKLNN-----NLVTFSS-----YNN 1085
Db      1463 LSPDTN--BQSLGLRLFEKRPD---QDTQENNVVQVNNPNNGDFLLPLLTASSQGPQLFSP 1517
QY      1086 FGALEPMVVFTAGSTLGIILAIMIILGLAIGIPLRAQRKLDKGFETTERKVDTLTAAGV 1145
Db      1518 FNQMPDYVLPLAITVPIVVIVLSVTGLAIGIPMHKNNKQALKAGFALSNQKVDVLTKAAG 1577
QY      1146 SVYKKIITQATANKKKRPAALGAGKSGDKRAAARAKPAAPASAPASAPAKTPGKSSGA 1205
Db      1578 SVKEIINRT--GISQAPKRL-----KQTSAAKGAAPRPVPAFGAKPPVQP--- 1624
QY      1206 PTKP 1209
Db      1625 PKKP 1628

```

```

RESULT 9
Q93005 ID 093005 PRELIMINARY: PRT: 1635 AA.
AC 093005:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Cytadhesin p1.
GN p1.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasma; Mycoplasma.
ON NCBI_taxid=2104;
RN [1]

```

```

RP      SEQUENCE FROM N.A.
RC STRAIN=MAC:
RX MEDLINE=21391823; PubMed=11500436;
RA Dorio-Zetsma J.W., Wilbrink B., Dankert J., Zaai S.A.J.;
RT "Mycoplasma pneumoniae p1 Type 1- and Type 2-Specific Sequences within
RL the p1 Cytadhesin Gene of Individual Strains."
DR EMBL: AF290001; AAK92039.1; -.
DR InterPro: IPR004940; Adhesin_p1.
DR Pfam: PF03257; Adhesin_p1.1.
SQ SEQUENCE 1635 AA; 176772 MW; 74FABB8D76091F53 CRC64;

Query Match 4.9%; Score 315.5; DB 2; Length 1635;
Best Local Similarity 20.4%; Pred. No. 6e-08;
Matches 324; Conservative 168; Mismatches 503; Indels 595; Gaps 82;

QY      41 LVNQARTLDANSYRLAGLGONGSLFNTLRDNDNFTTANGTTIKLDSFTKPLYGIDLS 100
Db      219 LPNQGSSSGSDST-----NOTGAMFGLKVD-----ATVDSKOSTESLKGESS 263
QY      101 DDCGGYKVKQIVSDYTTSRNRPOR-----OTRAYYA 132
Db      264 -----SSSTTSSTSTTQCGSSNENKVKALQVAKKSSGSGDQGTBOVE 311
QY      133 LTVNDEANVHLKRIINTNSRIGNRNNSKFIQVD--NPAHIVRFTDGTGFNFNTN-Q 188
Db      312 LESNDLANAPIKR-----GSNNQVQLAKADDFGTAPSSSGSGTQDGTPTPWTPL 362
QY      189 TQGEIYND-----FILAP-----ILPKDL----- 208
Db      363 TTEQIHNDPAKFAASILILVDAPYARRAIDRVHDLDPKAMTANYPPSWRTPKMNHGL 422
QY      209 -----HPDMYN--LYIORKI--LPPDVNT----- 228
Db      423 WDKKARDVLIQTGFENPRRHPMEFDGGQTVADNEKTEFDVDNSENTKQFQKEADSDKS 482
QY      229 AVVPMP-----VGRVSGTNADGMEFCGNGQIT--NTDPIA-----QTKTT 267
Db      483 APIALPFEAFANIGNLTWEGQALLVFG--GNGVHTSAHAPLSIGFVRYNATGTSAT 541
QY      268 TDQNDPSTNSGMPGANNRYDSQLNVKRIKTSFQDERINTNSRIGNRNNSKFIQVIG 327
Db      542 VTGMPYALLFSGAV---NKQTDGLKMLPENNRMFEVPPMAVAGAKFVGR---ELVLA 594
QY      328 G---VONPAHIVRFTD--GTFKFNTOGELVNPIL-----DAPT----- 365
Db      595 GTITMGDTATVPRLYLDELESNLNLAQGGGLREDLQLTPTYGMANRPDLPIGAMSSSS 654
QY      366 -----LPKDLH--PDWYNTLYIORKILPNDVNTAVVPVGRVSGTNAD-----G 408
Db      655 SSSHNAPYVHHNNPDMQDRIO-----SVVDARLIKWE---DKNGDKAKYIYPYRSG 705
QY      409 MEDCGNGQITN--TDPIAOKTKTTTD--NONPSTNS-----GAMP-----GANN- 448
Db      706 MW---AWQVYVNMENKLTDPQLSADFVENAYQFNSJLFAAILNPELLAALDPKAYKGENE 762
QY      449 -----RYDSQNLVK-----HRTKTSPO-----LDEKFVYPEWYSGENK 482
Db      763 PAANEYERENOKLTVAPTOGCTNMSHFSPTLSRSTGBNLVGSVLDQVLPVWIGNGYR 822
QY      483 NITRLATGSLPSNERWIIDIPCTPOVTLKEDSVNFSRLYLSVNSLSFI----- 533
Db      823 G-----NNHR-----GVDDITAPQTSAGSSSGISTMTSGSRSLPFFSNIGVG 865
QY      534 -----GDSIYIFGISELPSLWYSPPPLSLDITLALNOYKT---DIEASSTNG 579
Db      866 LKANVOATLGGSGQTMITGGS-----PRTLLDQANL-OLMTGAGWRNDRKSSGSD 914
QY      580 TTNGTITTTADTSSGSGAGTGN-----TTNNSQVSVNPTLWTVRSFGI 623
Db      915 DHTKFTSATGMDQGGSGISAGNPDSLKODKISKSGDSLTTQDGNATGGOEATVYTNLPP 974
QY      624 DSKPT-----SANKIDETNW----- 638

```

Db	1034	1033	1032	1031	1030	1029	1028	1027	1026	1025	1024	1023	1022	1021	1020	1019	1018	1017	1016	1015	1014	1013	1012	1011	1010	1009	1008	1007	1006	1005	1004	1003	1002	1001	1000	999	998	997	996	995	994	993	992	991	990	989	988	987	986	985	984	983	982	981	980	979	978	977	976	975	974	973	972	971	970	969	968	967	966	965	964	963	962	961	960	959	958	957	956	955	954	953	952	951	950	949	948	947	946	945	944	943	942	941	940	939	938	937	936	935	934	933	932	931	930	929	928	927	926	925	924	923	922	921	920	919	918	917	916	915	914	913	912	911	910	909	908	907	906	905	904	903	902	901	900	899	898	897	896	895	894	893	892	891	890	889	888	887	886	885	884	883	882	881	880	879	878	877	876	875	874	873	872	871	870	869	868	867	866	865	864	863	862	861	860	859	858	857	856	855	854	853	852	851	850	849	848	847	846	845	844	843	842	841	840	839	838	837	836	835	834	833	832	831	830	829	828	827	826	825	824	823	822	821	820	819	818	817	816	815	814	813	812	811	810	809	808	807	806	805	804	803	802	801	800	799	798	797	796	795	794	793	792	791	790	789	788	787	786	785	784	783	782	781	780	779	778	777	776	775	774	773	772	771	770	769	768	767	766	765	764	763	762	761	760	759	758	757	756	755	754	753	752	751	750	749	748	747	746	745	744	743	742	741	740	739	738	737	736	735	734	733	732	731	730	729	728	727	726	725	724	723	722	721	720	719	718	717	716	715	714	713	712	711	710	709	708	707	706	705	704	703	702	701	700	699	698	697	696	695	694	693	692	691	690	689	688	687	686	685	684	683	682	681	680	679	678	677	676	675	674	673	672	671	670	669	668	667	666	665	664	663	662	661	660	659	658	657	656	655	654	653	652	651	650	649	648	647	646	645	644	643	642	641	640	639	638	637	636	635	634	633	632	631	630	629	628	627	626	625	624	623	622	621	620	619	618	617	616	615	614	613	612	611	610	609	608	607	606	605	604	603	602	601	600	599	598	597	596	595	594	593	592	591	590	589	588	587	586	585	584	583	582	581	580	579	578	577	576	575	574	573	572	571	570	569	568	567	566	565	564	563	562	561	560	559	558	557	556	555	554	553	552	551	550	549	548	547	546	545	544	543	542	541	540	539	538	537	536	535	534	533	532	531	530	5
----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

DR	InterPro: IPR004940; Adhesin Pfam: PF03257; Adhesin Pfam: PF03257; Adhesin Pfam: PF03257	Score 311.5; DB 2; Length 1635;
DR	SEQUENCE 1635 AA; 176888 MW; F2D3795740D5607E CRC64;	
Query Match	4.9%;	
Best Local Similarity	20.4%;	Pred. No. 9,6e-08;
Matches	325; Conservative 167; Mismatched 503; Indels 595; Gaps 82;	
QY	41 LVNQAATLDDNSVRILAGLGNGSLFETVLRDVEDNFTTANGCTIIKIDSTFKPLTGIDL 100	
DB	219 LPNGSSSGSDST---NOTGAMFLAKYD-----ATVDSKOSTESTIKGEES 263	
QY	101 DDCGKYKQIVSDYTTSSNRFDOR-----OTRAYA 132	
DB	264 -----SSSTTSSTTQKGSSNENKAYALOVAAKKSGSGNGDGTQVE 311	
QY	133 LLYNDEANVHLKRIINTSNIRIGNNNNSKFEIVGVD--NPAHVRITFDGCTKFNFN-Q 188	
DB	312 LESNDLANAPIKR-----GNNNOVOLKADDFGTAPSSSGSGTQDTPTPWIPWL 362	
QY	189 TQGEIVND-----FILDAP-----ILPKDL----- 208	
DB	363 TTEQIHNDFAPKASILITLDYAPARRKRTAIDVHDLDPKAMTANYPSPWTPKNNHGL 422	
QY	209 -----HPDWN-----LYIQRI-----LPDVT----- 228	
DB	423 WDKKADVLLQTTGFENRHRPEWFDGGQYVADNEKTEGPDVDSNTKQGFQKEADSDKS 482	
QY	229 AVYWP-----VRSVGTNADDMFDCNGQIT---NTDPIA-----QTKT 267	
DB	483 APTALDFEAYIFANIGNLTFQGLLVFG-GNGHYTKSAHTAPLSIGFVRYNATGTSAT 541	
QY	268 TDQNSSTNSGAMPANRRVDSQNLVNRKIKTSFQDLRIINNSNRIGNNNNSKFEIVG 327	
DB	542 VTGMPALLFSGMW---NKQTDGLKNLPPNNNRFEVYPRRAVAAKFWGR---ELVLA 594	
QY	328 G---VONPAHVRITD---GTFKNFTNTOGEIYNDPIL-----DAPL----- 365	
DB	595 GTTMDGTALVPRLLXDELESMLNVAQOGLIREDLQTFPYGMANRPDLTGAMSSS 654	
QY	366 ----LPRDLH---PDWYNLYIQRIKLPPDYNVNAVVPMPGVSGSTNAD-----G 408	
DB	655 SSHNAYFYHNPDWDQDPIQ-----SVDAFIRPW-----DNKGDDAKYIYPYRSG 705	
QY	409 MFDGNGOINR-TDPIAQKTTD--NQNPSTFN-----GMP-----GANN- 448	
DB	706 MM---ANQVYVMSKKLIDQLPSADFVFNENAYPNISLFAAILNPELLALPKVYKGENE 762	
QY	449 -----RYDSQLNVK-----HRIKTSFQ-----LDEKFVPEWMTGSEENK 482	
DB	763 FAANEYERFNOKLIVAPDTGNTMSHSPILSRSTGFENLVGSVLDOVLDYPMVINGRY 822	
QY	483 NTRLATLSLPSNEREYWLIDPTGPOVTEKEDSVNFSLYLNSVNSLSFL----- 533	
DB	823 G-----NNHR-----GYDDITTAQTGAGSSSGISNTGSHSLPTPSNIGV 865	
QY	534 -----GDSIYFCTSELPFLMWYSEPTSLSDLTFLINQYTD---DIEASITDG 579	
DB	866 LKANYAQTLLGSGQMITGS-----PRRTLDQANL-QLMTGAGWRNDKASSGSD 914	
QY	580 TTTNGTNTTADTSAGSGTAGTGN-----TTNNSQVSNVPLNTYRSGFI 623	
DB	915 DHTKFLPSAAGMGQOQSGTSAGNDPSLKDOKISKSGDSLTTQDGNAMDEQDEATNTNLP 974	
QY	624 DSKPT-----SANKIDETW----- 638	
DB	975 NLPTPTADMPNALSFTKNNNAQAQLFLKGLGSLPVLNKSQGDNDNSKFAEDQKSYND 1033	
QY	639 --ADPNIIEARLYAEYRIGIONEIP-----TJAGNFRINTG--GVGF-----NST 681	
DB	1035 LOSDQKFLNIPAYGEVN-GILNPAIVETVFGNTRASSGSGSTTSSPIGKRIEOSTGNT 109;	
QY	682 GSRKIVL-RASYNDQPTGN-----FQ--PFLYF-----GYLQY----- 714	

[illegible][illegible]

[illegible]

Db	SLNTERVEIN	YTEKLNENETQKSTYITKNNNGKWTINKKNYVEFN	1689
Qy	975 ONLSDIAFEESGAKYTSDFWCTIQFKPDEYLIQNGFTSOVARNFVTNOSFLNSLVDETPA	1034	
Db	1690 ODNGKVFVFSATIKPNSQI--TIRPKAGQGNTEN-----TNPVYIAQ-----PA	1721	
Qy	1035 NAGTNIRKVVYDDPG--NLTNQMLPLKVOIQYIDGKYYDAKLNKNNLVYFSTYNFCALPSPW	1093	
Db	1732 QHTLTINEIVEKEQONVYNDINDINNAVQ-----PNKNRVAIKQGN--ALPP--	1775	
Qy	1094 VPTAIGSTGLIATM 1108		
Db	1776 -NLGGSTSHIPVYI 1789		
RESULT 14			
ID	08GR92	PRELIMINARY;	PRF; 4727 AA.
AC	08GR92:		
DT	01-MAR-2003 (TREMBLrel. 23. Created)		
DT	01-MAR-2003 (TREMBLrel. 23. Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23. Last annotation update)		
DE	GL1521.		
GN	GL1521.		
OS	Mycoplasma mobile.		
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.		
OX	NCBI_TaxID=2118;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Seto S., Uenoyama A., Miyata M.;		
RT	"Identification of GL1521, Force Generating Protein, for Gliding		
RT	Motility of Mycoplasma mobile."		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB084781; BAC23069.1; --		
SO	SEQUENCE 4727 AA; 520559 MW; 24BEC67133DBD2EC CRC64;		
Query Match 3.9%; Score 250.5; DB 2; Length 4727;			
Best Local Similarity 20.8%; Pred. NO. 0.00059;			
Matches 282; Conservative 176; Mismatches 487; Indels 409; Gaps 71;			
Qy	19 FGALGSASFGFKOSDKSDNDNTQLVNQAFTLDANSVRLAGL-----QNGSLFTVLRDVD	73	
Db	1622 FETLKSTSVIIGGSDSKN-----LSINFGLSLSSLRSGIDVQFSINNGTWESYL--VE	1675	
Qy	74 DNFITAA---NGTIIKIDSPFKRLPLXGLDLSDDCGYKKKQIVSDYTTSRNFDQRQTPAY	130	
Db	1676 GOLIDASVNDPTKVSLSLRLPRNNEPAREISIQ-----	1707	
Qy	131 YALLV-NDENAVHLKRIPTNSNR---IGNRNNNSKFVIGVDNPAHYIRFTDDGKFNFT	186	
Db	1708 -ALVYKYNQAKILISGLINNNLRDANLVNLSGNTRDLAINVQ--NVRKATGNSPTTPT	1764	
Qy	187 NOTQGEIYNDLIDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVPMPEVGRVSGTNADG	246	
Db	1765 NIIVQRFALINGVLGGTFPEKD-----EFLRELNSLPKSIINAIIRPCDF--HVSFTLLNNE	1814	
Qy	247 MEDCGNG-----QINTDPIAQTQKTTDDNQNSTFNSGAMPANNRXYD--SOLNVKHRITS	301	
Db	1818 DFLSEEGISTQPIPIPTIVNKKRLVLDLHDVNLANGSSIVSSGSTSLSLSNPSSLSETTL	1877	
Qy	302 FOLDERINTNSRIENRNNNSKFVIGVDN---PAHYIRFTDDGKFNFTNOTQGEIYND	358	
Db	1878 TOLGIVLEKGSYR--RANGKIDVNVYWEHTIPRAISATPRDRINQNF-----	1925	
Qy	359 FILDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVPMPEVGRVSGTN-----ADGMDPG	413	
Db	1926 -----PAPINSAFSLAFRAFKII--DEMTTII-----VNGNSINIVNCGVD--DLG	1968	
Qy	414 NGQITNT---PPIAQTKTTDDNQNSTFNSGAMPDA-----NNRYDS--QLNV	456	
Db	1969 PSINNTSLKRIIDFTTTTSMFINNDLQKIVLGTTNTINQOLISHSENNLNTSIITQDI	2028	

NCBI:taxid:7027

SEQUENCE FROM N.A. cypnus;
STRAIN=Kenya tick cypnus;
MEDLINE=93194085; PubMed=7680636;
Gillmore R.D. Jr.;
"Comparison of the rompa gene repeat regions of Rickettsiae reveals
species-specific arrangements of individual repeating units";
RL Gene 125:97-102(1993).
DR EMBL, 101462; AAA9908.1; -
FT NON_TER 1 1029
FT NON_TER 1029
SQ SEQUENCE 1029 AA; 99999 MW; 09573881A5B9BDCL CRC64;

Query Match 3.9%; Score 247.5; DB 2; Length 1029;
Best Local Similarity 20.9%; Pred. No. 0.00011;
Matches 269; Conservative 131; Mismatches 468; Indels 417; Gaps 62;

7 LKSYLLIGLAVF-GALGSASFGSKDSKSNQTQLVQACATLDANSYRLAGIQNGSLF 43
7 LRTVVGAGIATTEGAIKAT-----TTKLTAASVLTLTNV----- 43
66 NTVLEVDVNDNTTAANGTIILKDSFTKPLYG-----LDSDDCGYKVKQIVSDYTT 117
44 NNVTLGALIDNTTGVNDNGVLTNLGALSQVTGNGIGNATLATTISVGAGATLGAVIKAT 103
118 SNRPDQROTIRAYALLY-----NDEAN---VHLKRTNSNRIGNRN-----NSK 161
104 TKLTLDASQVFTNPVVVTGALIDMTGANNNGIYTFGDSVTYVGNIGNATLATTISVGAG 163
162 FVLGVNDPAHVIREFTDGTRKFNFTNO--TQGEIVNDFILDAPILPKDLHPDWNLYIQ 219
164 ATLGAIKATTKTKLTLDASQVFTNPVVVTGALD----- 199
220 KILPDVNTAVVWMP-----VGRVSGTNADDMGFCNGQITNTDPIAQTKT--TDNON 272
200 --TGANNNGIYTFGDSVTYVGNIGNATLATTISVGAGATLGAIKATTKTKLTLDAS 255
273 PSTP-NSGAMPAGANNRYDSQLVKHKIKTSFOLERITMSNRIGNRN-----NSKF 324
256 QVFTNPVVVTGA--IDNTGANNNGIYTFGDSVTYVGN--IGNATLATTISVGAGKA 308
325 VVGVDNPAHVIREFTDGTRKFNFTNO--TQGEIVNDFILDAPILPKDLHPDWNLYIQ 382
309 TLGGAIKATTKTKLTLDASQVFTNPVVVTGALD----- 343
383 ILPDVNTAVVWMP-----VGRVSGTNADDMGFCNGQITNTDPIAQTKT--TDNONP 435
344 --TGANNNGIYTFGDSVTYVGNIGNATLATTISVGAGATLGAIKATTKTKLTLDASQ 400
436 STF-NSGAMPAGANNRYDSQLVKHKIKTSFOLDEKFEVPEWTGSE--ENKNITRLATG 490
401 VFTNPVVVTGA--IDNTGANNNGIYTFGDSVTYVGNIGNATLATTISVGAGATLG 445
491 SLPSNRERYILDIPTGPVLTLEKDSVNVFSR--LYLNSV-----NSLSF 532
446 NVAG--LKVQGG--VKSNTINLTLDASQVFTNPVVVTGALIDMTGANNNGIYTF 498
533 IGDSTIYI-----FQSELSPLMYISFPTRLSD-----LITALNGYKXDDIEA 573
499 TGSITYVGNIGNATLATTISVGAGATLGAIKATTKTKLTLDASQVFTNPV--VYT 554
574 SSTDN-GTTTNGTTTADTSSGSTGAGCIGNTNTSQTYSNFTLNTYVSFGIDSKPTSA 632
555 GADINTGANNNGIYTF--FTGGSVTYVGNIGNATLATTISVGAGATLGAVKVSNTINL 610
633 IDEINNADPNVITARIYAERLGIQNEIPITNAGNFINTTGGVGTSTGSRVLRASYN 692
611 TDNAS--I-----QVFTNPVVVTGALIDMTGANNNGI--VTF 643
693 GDSPTGNFQFLVYVGLGYQOTRGCTEMYGYYKLANSPYDADSPVGTETNOGRTT 752
644 GDSVTGN-----IG-NTNAL-- 658

```

Oy 753 SLTPWAGGYEEGASFSWTPYIRAGDPE---SRSLFGSVDNTXEYIGSYVLGFD 809
Db 659 -ATVWAGGLLRVGCGVYKST--INLTDAISOYTFINPVVYTGADNT----- 704
Oy 810 GIRNNLVGKASSFLNSRPNGLLEMI--AATYLRSGIAGTSGLPNOQPGCTH 866
Db 705 GNANNGIYTFEGNSTVGNIGNTNALTATVWAGIATLEGAVIKATTTKLTINASVTLT 764
Oy 867 QVISVSPC--DQFSIKNIRTIIFGNOIMYELFINENKSKSVYTLRLADSNPSSSFS 924
Db 765 NNAVALTGADINTTGVDNAGVL-----NLGALISOYV--GNIGNTNALTATIS 809
Oy 925 -----PTSLID-----VNEIGVILPDLDSFYETNAGANVALF---S 958
Db 810 VGAKATILGGAIVIKATTTKLTLDNASOYTFINPV-VVIGALDN---TGNANNGIATFTGDS 865
Oy 959 SNPSPGSGYIAVNTFFNONLSDIAEFGSGAKYTSDFMGTIOFKPDEYILONGFTSOVARNF 1018
Db 866 TVTGNIGNTNALTATVN-----VGAGL-----LRVOGVVYKSTINIL 901
Oy 1019 VTNOSFLSLVDFTPANAGTNYRVVYDPDGNLTQNQLPLKQIQYLDGKYDADKLKNNNL 1078
Db 902 TDNA-----SAVTFINPVVYTG--AIDNTGN-----ANNGI 930
Oy 1079 VTESYNNNEGALPWSVVPALGSTGLIAMIILGLAI-GIPLRA-ORKLDDKGFKTFK 1135
Db 931 VTFTGDS-----TVTGNIGNTNALTATISVGAKATILGAIIKATTTKLTLDNASOYTFET 983
Oy 1136 KVDFTLTAAY--GSVYKKIITQTN 1157
Db 984 NPVVVYTGADINTGNANNGIYTFETD 1008

```

Search completed: October 10, 2003, 16:35:18
Job time : 89 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 13:34:27 : Search time 21 Seconds

(without alignments)
2740.985 Million cell updates/sec

Title: US-09-901-572A-4

Perfect score: 6413

Sequence: 1 MNISKKLSYTLIGSLAVFG.....APTKPPAPKAPKPTAPKE 1224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598.5	9.3	1052	1 MGPC_MYCGE	P22747 mycoplasma
2	556.5	8.7	1218	1 MGPC_MYCPN	Q50341 mycoplasma
3	501.5	7.8	1122	1 ADPL_MYCGA	Q49379 mycoplasma
4	328	5.1	1627	1 ADPL_MYCPN	P11311 mycoplasma
5	285	4.4	1444	1 ADPL_MYCGE	P20796 mycoplasma
6	232.5	3.6	1421	1 OMPA_RICCN	Q52857 rickettsia
7	216.5	3.4	1643	1 OMPA_RICCP	Q53020 r outer mem
8	215	3.4	1656	1 OMPA_RICCN	Q06653 r outer mem
9	204	3.2	1655	1 OMPA_RICCN	Q06653 r outer mem
10	202	3.1	1378	1 YS89_CAEEL	Q09624 caenorhabdit
11	199	3.1	1256	1 MRP_SPRSU	P32653 streptococc
12	189	3.1	1902	1 P3P_LACLC	P15282 lactococcus
13	198.5	3.1	2249	1 OMPA_RICRI	P15921 rickettsia
14	197	3.1	1306	1 MSB2_YEAST	P32334 saccharomyc
15	192	3.0	1300	1 120K_RICRI	P14314 rickettsia
16	192	3.0	1654	1 OMPA_RICRI	Q53047 r outer mem
17	190	3.0	2710	1 TOXA_CLODI	P16154 clostridium
18	189.5	3.0	1645	1 OMPA_RICRI	P36666 escherichia
19	189.5	3.0	2003	1 YDBA_ECOLI	P33666 escherichia
20	187.5	2.9	1288	1 VACA_HELPJ	Q92KW5 helicobacte
21	187	2.9	1902	1 P2P_LACLC	P15293 lactococcus
22	186	2.9	1902	1 P1P_LACLC	P16271 lactococcus
23	184	2.9	1861	1 APU_THETU	P38536 t amylolipul
24	182	2.8	1902	1 P2P_LACPA	Q02470 lactobacilli
25	179	2.8	1723	1 PM20_CHLPP	Q92812 chlamydia p
26	175.5	2.7	1858	1 P3K2_DICDI	P54674 dictyostell
27	175	2.7	1672	1 PMPB_CHLMD	Q09132 chlamydia m
28	174	2.7	2334	1 WAPA_BACSU	Q07833 bacillus su
29	172.5	2.7	1287	1 BICA_HELPY	Q48245 helicobacte
30	172.5	2.7	1953	1 BICA_SALT	P25927 salmonella
31	172	2.7	1802	1 HKRI_YEAST	P41809 saccharomyc
32	172	2.7	1849	1 IGAA_HAEIN	P45386 haemophilus
33	170.5	2.7	833	1 HSF_YEAST	P10961 saccharomyc

34	167.5	2.6	1076	1 NUPL_YEAST	P20676 saccharomyc
35	166	2.6	1256	1 ATL_STRAU	P52081 staphylococ
36	165.5	2.6	1291	1 VACA_HELPY	Q48258 helicobacte
37	165.5	2.6	1778	1 N189_SCHPO	Q94K4 schizosacch
38	165	2.6	1075	1 FLOS_YEAST	P38894 saccharomyc
39	165	2.6	1848	1 CBPA_CLOCL	P38058 clostridium
40	164.5	2.6	1464	1 NC02_HUMAN	Q15596 homo sapien
41	164.5	2.6	2358	1 YEEL_ECOLI	P76347 escherichia
42	164	2.6	1325	1 YDEK_ECOLI	P32051 escherichia
43	163	2.5	1228	1 SLAP_BACST	P35825 bacillus st
44	163	2.5	1419	1 ALAI_CANAL	Q13368 candida alb
45	163	2.5	1537	1 FLOI_YEAST	P32768 saccharomyc

ALIGNMENTS

RESULT 1
MGPC_MYCGE STANDARD: PRT: 1052 AA.
AC P22747: Q49257: Q49284: Q49385: Q49481: Q49482;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mgpc protein precursor.
GN MGPC OR MG192.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=90060815; PubMed=2583522;
RA "Inamine J.M., Loechel S., Collier A.M., Barile M.F., Hu P.-C.;
RT "Nucleotide sequence of the Mgpc (mgp) operon of Mycoplasma
RT genitalium and comparison to the Pl (mmp) operon of Mycoplasma
RT pneumoniae.";
RL Gene 82:259-267(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RL "The minimal gene complement of Mycoplasma genitalium.";
RN Science 270:397-403(1995).
RN [3]
RP SEQUENCE OF 60-133, 260-370, 441-512 AND 964-1052.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
RN [4]
RP SEQUENCE OF 769-964 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=92051396; PubMed=1945886;
RA Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A random sequencing approach for placing markers on the physical map
RT of Mycoplasma genitalium.";
RL Nucleic Acids Res. 19:6027-6031(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@slb.ch).

CC -----

DR EMBL: M3131; AA25421.1; -

DR EMBL: U39698; AAC71411.1; -

DR EMBL: U02124; AAD12399.1; -

DR EMBL: U02161; AAD12443.1; -

DR EMBL: U34967; AAA88890.1; -

DR EMBL: U02157; AAD12439.1; -

DR EMBL: U34970; AAA88899.1; -

DR EMBL: X61525; CAB98130.1; -

DR EMBL: X61327; CAB98131.1; -

DR PIR: C64221; C64221. -

DR TIGR: M6192; -

DR pfam: PF05220; MGPC; 1.

KW Cell adhesion; signal; Membrane; Complete proteome.

FT CHAIN 1 25 POTENTIAL.

FT CONFLICT 26 1052 MGPC PROTEIN.

FT CONFLICT 125 126 LDSSYQAD -> MIKPLPLLS (IN REF. 3).

FT CONFLICT 474 481 TM -> SV (IN REF. 3).

FT CONFLICT 476 476 TNNGTGP -> SNQISSGT (IN REF. 3).

FT CONFLICT 476 476 N -> S (IN REF. 3; AAA88890).

FT CONFLICT 897 902 SVSPRT -> MSHGCS (IN REF. 4).

FT CONFLICT 1052 AA; 114360 MW; F6C6D3BD7789145 CRC64;

SO SEQUENCE

Query Match 9.3%; Score 598.5; DB 1; Length 1052;

Best Local Similarity 23.7%; Pred. No. 9.8e-24;

Matches 270; Conservative 166; Mismatches 409; Indels 295; Gaps 49;

QY 260 PIAOTKT---TTDNQNSTFNFGAMGANNRDSOLNVRKIKTSFOLDERTINTNSN--- 313

DB 21 PLALNTFLVKEDSKNTATYTPATPTITDSK--SDLVSLQADSSQIADQTHNTLNFV 78

QY 314 RIGRRNNNSKFKVIGVDN-----PAHVIRETDGTFKNEFTN--OTQGEIYNDFL 361

DB 79 LFKSDVAVKVESSGNNISFSDTSQEKPSYVEFT-----NSTNIGIKMTWKKKYQL 132

QY 362 DAPILPKDLHPDMYNYLQKILPNDVNTAVWPVPRGVSCTNADGMPDCNGQITNTD 421

DB 133 DVPNVSSDMQVLNKLLEOPLTKYTLNLS-----LAKKGGKQRE--VHLGSG----- 179

QY 422 PIAOTKT---TTDNQNSTFNFGAMGANNRDSOLNVRKIKTSFOLDERTINTNSN--- 313

DB 180 ---QANQMTSQRNQNDLNNPNSPANSNGFKLTGNARVRLKLSSEPTIEPTIDGTQKQKMD 236

QY 475 ---WTGSEEN--KNITRLATGSLPSN-----ERYWLDIPGTPQVTLKEDSV 516

DB 237 SSGWSTEBENKAKNDAPSVSGSSGSGTFNKYLTWKQALBSIGILFDDQTPR----- 288

QY 517 NVFSRLYLNSVNSLSFIDGSIYIFGTSELPSLWYSEPTRLSDLTALNQKTTDIEASST 576

DB 289 NVITQLYASTSKLAVTNHILYVGNNSFLPSMWYV-----VERSQAQ 330

QY 577 DNGTTNGTTTADISSG-----STGAGTNTNT-----SOTVSNPLNTLYRS 620

DB 331 EN--ASNKPTEFANLMDGEDKOKQFVENOLGYETTSNHNHSSFPQF---AVLI 385

QY 621 FGID-----SKPTSAKIDETNADPNVTEARIVAEYRIGI 656

DB 386 SGIDSVNDQITIFSGFAAGSVGYDSSSSSSSSSTRKDALASTTSSDSK--TCYKDLV 443

QY 657 QNEIPIITNAGNFIKRTIGVGTSTGSRVYLKASTNGDORPFGNQPFLYVGVGYGYOQT 716

DB 444 TNDGANGPIN-----GSFSIODTFSPFV---PYSGNHTNNGTGPITKATAYPVAKDKS 494

QY 717 RTGFWYGYK---LLNNSPYDVLDSPRGT-----ETNOFRRTSLTPYV-MGQY 762

DB 495 -----YKINSLINATPLNSTYGDGIEVPALGLANFNKSNQERLSPRDOJFYVGI 546

QY 763 LTBEGARSEWTPYIRAGDTPESRSIFQSGYSDNTYEQSYLVGDLGIRNNLNVGARS 822

DB 547 VSPNELRS-AKSSADSTSGDTKVMWMTQSRYLFPVYNSSEGIIDADGKRPENRGASVT 605

QY 823 SFLNSKPNPNCLEMTAATTYLS-----OIGLARTSGL-PNQOPFG--THQVIVSVSPGD 875

DB 606 TF-----SGLKSLAPDGFANSTANFSVGL--KAGIDPNVMGSKRANGAVLTFRG 655

QY 876 QFSIKNIRTF--PGNOLWYFLFTNENN-----KS 904

DB 656 V-----VRLNPNNGND--SLSTTDNNIAPISFSEPTPTAESAVALDTTEEVYNOES 707

QY 905 SVYTLRLADSNPDASSFSPT-----SLIDVNEIGVILLPLDINSFYTVNAGN-VALF 957

DB 708 GLMSYIFDSSLSKPSHDKQTPVDNMGFSYITYSRFGIELN--QDAFTTLIDVAFPAALAVQ 766

QY 958 SSNPGSPGSYTAV-----NTEPNLSDI-----AF 985

DB 767 SGIGSTQTLGLVPLSEFEFSAVIAKDSQNKIDIKNNNGLEIDITQLSNVSATNGL 826

QY 986 GAKYTS---DWGTLQFPPDEYLLQNGFTSGVARNFYTNOSFLNSLVDTFPAAGTNYV 1042

DB 827 APSYENRYDAMGKVEFADNSVLAARNLVDKTVDEILNTEPELINSFREFTPAEDOKAVL 886

QY 1043 VYDPDGNLTNQLPLKVOIOYLDGKYVDKLNKNNLVTFSYN-----NFGALPSWY 1093

DB 887 VATKQ---SDTSLVSPIQFLDGNFYDL--NSTIAGVPLNIGFSPRVRPAGFALPAMV 940

QY 1094 VPPAIGSTIGILAINITIGLAIPIRAQRKLDQKFTTKKVDLTAAVGSYKKIIT 1153

DB 941 IPVSAGSSGILFVLVGLGIGIPMYVRKLDQASFNVRKRVDTLTAAVGSYKKIIT 1000

QY 1154 QFANVKKRPAALGAGSGDKRPAAPAKPAAPKAPASSPAKPTGPKSGAPKPAAPK 1213

DB 1001 QTVGVKKAPSALKAAPSVKKAAPLKP--PVQP-----PSKPEGBOKAVEKSEBTK 1051

RESULT 2

MGPC_MYCPN STANDARD; PRT; 1218 AA.

ID MGPC_MYCPN 050341;

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE MGPC protein precursor

GN MGPC OR MPN142 OR MP012.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [11]

RN SEQUENCE FROM N.A.

RP STRAIN=ATCC 29342 / M129;

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97103865; PubMed=8948633;

RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

RA Hermann R.;

RT "complete sequence analysis of the genome of the bacterium Mycoplasma

RL pneumoniae";

RL Nucleic Acids Res. 24:4420-4449(1996).

RL [31]

RP IDENTIFICATION BY MASS SPECTROMETRY.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=21088919; PubMed=11271496;

RA Reguana R., Ueberle B., Boguth G., Goerg A., Schnoelzer M.,

RA Herrmann R., Frank R.;

RT "Towards a two-dimensional proteome map of Mycoplasma pneumoniae";

RL Electrophoresis 21:3765-3780(2000).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, M21519; AAA83326.1; -
 DR EMBL, AE000002; AAB95660.1; -
 DR PIR, J50069; J50069.
 DR Pfam, PF05220; MGPC, 1.
 KW Cell adhesion; Signal; Membrane; Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1218 MGPC PROTEIN.
 SQ SEQUENCE 1218 AA; 130456 MW; 4DA29BCE41538311 CRC64;

Query Match 8.7%; Score 556.5; DB 1; Length 1218;
 Best Local Similarity 22.1%; Pred. No. 1.8e-21;
 Matches 321; Conservative 176; Mismatches 483; Indels 473; Gaps 65;

QY 1 MNISKLKSYTLIGLAVGALGSASFQKSDKNDNTOLVNOARTLDANSVRLAGLQ 60
 DB 1 MKSKLKLRYLLFLPLPLGLTSLA-----NTYLLQDHNITLTPY----- 40
 QY 61 NGSLENTVLRDYNDFITANGTI- IKLDSFTKPLYGLDLSDCGQYKQIVSYTTSR 119
 DB 41 -----FTTPPLNGCLDVVRAHLHPSEYELVDMKRVGDTKLVALVRS-ALVR 85
 QY 120 NREDFORIRAYALLVNDENAVHLKRINTNSNRIGNRNNNSKFYIGVDNPAHVIRFTD 179
 DB 86 VFQDITSS-----SDQS-----NTNQLASPDIOESQKALNG-----SQS 120
 QY 180 GTFNFNTOTGEIYVNDFFILDPILPKDLHPDYNLYIQRLKLPNDVNTAVVWPVGRYS 239
 DB 121 G-----SSPTSSNSQDFASVYLIFKAPRATW--VEPKRI-----KLALPY-VKQSS 165
 QY 240 GTNADGMFDCNGOITNTDPIAQRTT--TDN-----QNPSTENGAMPGANR 287
 DB 166 QSSGQGSNGKSLKTTLODLLEQVPVPTYNAGLARVNGVADTVHFSGQSSSMNSQ 225
 QY 288 YSOLNVK-----RIKTSQFLDERINTNSNRIGNRNNNSKFYIGVDNPAHVIRFTD 343
 DB 226 -RSQGLKNNPGPKAVTGFKLCK-----GRAVRLKLNESWPYEPDLDS----- 266
 QY 344 KENFNTOGEIYVNDFFILDPILPKDLHPDYNLYIQRLKLPNDVNTAVVWPVGRYS 402
 DB 267 -TKEGK-----KD-ESSMKN-----SEKTLAENDAPLYGVMS 298
 QY 403 TNADGMFDCNG-----QITNTDPIAQRTTNTDNPSTFN----- 439
 DB 299 GAAGSASSLOGNSNGSLKSLRSAPVSVPPSTNSQTLISNPAPVGPQAVVSOQAG 358
 QY 440 SGAMPANNRYSQNLVKKRITKTSFQDLCKFY-----YPMWTSSEKKNTIRLATGSLPN 495
 DB 339 ARAAVSVNTASDTATFSKYLTNTAQLHOMGVIVPGLKMWGNN-----GTGVVSR 410
 QY 496 ERYWILDP-----GTPQ-----VTL-----KEDSV-NVVS 520
 DB 411 QDATSNLPHAGASQGTGLGTSPPRALTATSQRAVTVYVAGPLRAGNSSEFDALPNVIT 470
 QY 521 RLYLNSVNSLSTIGSITYEGTSELPSLWY-----SEPTRLSDLT 561
 DB 471 QLYHTSTQALAYLNGOIVMGSDRVPSTLWYVWVGDEQSGKATWMAKTELNMGTDKOKOF 530
 QY 562 ALNOKV-TDDIASSSTNDGTTNGTTTADTSSG-----STGA-GTGTNTT 606
 DB 531 VENQIGFKDDSDNSKSLKAQGLQPPYLLAGLDVADHLVFAAFKAAGVDMTTS 590
 QY 607 SOTVSNPTLNTYRSGFIDSKPTSAKIDETNNADPNVIEARLYAEYRLGIONEIPITNG 666
 DB 591 SASSTYNQALAMSTTAGLSD-----GGKALVENTAGLNGI 627
 QY 667 N-----FIRNTIGGVFTSTGSRVYLRASTNGDORPPNGFOPPLVFEYGLGYQT 716

DB 628 NGLFTLLDTFAVVTVPVSGMKGGSONNEEYQTTYPVSKDOKATAKIA----- 673
 QY 717 RRGTFWYGYTKLNNSP-----YDV-----LDSRPGTETNOPRRSLPYP 757
 DB 674 -----SLINASPPLNSYDDGVYVVDALGLNFKLNERNLPSRTDL-----LVY- 718
 QY 758 VMGYLVE-----EGARSPSN-----TPYIRAGDPTES-----RS 788
 DB 719 ---GIYNESELKSARRNAQSTSDDNSNTKVKYNTNASHLYLPYPYYS-ANPEAGRRRA 774
 QY 789 IFOSGYSDNTYE-----YIOSVLFQDITRNMLNGVKASSFLNSRPNPNCLEMIATP 842
 DB 775 EDRNGVKISTLESQATDGFANSLNNG--TGLKAGVDPAAPARGHKPYSVALLVRG-- 829
 QY 843 YLRSGIARLARTSGLPNQ-----PFGTTHOVISVSGDGFSSIKNR 884
 DB 830 -----GVVRLNENPDTKLLDSTDKNSPISPSYTPFESASAVDL-----TTLKDV- 876
 QY 885 TIFPGNQLWYFLFTN-----ENKS-SVYTL-RLADSSNPASS--SFSEP 926
 DB 877 TYIAESGLMFYTFDNGEKRTYDGKQOVKRRGVAIIVYSKRGIEFNEDANTTLLSQAPA 936
 QY 927 SLIDNEIGV-----ILPLDNEFTYVNA-----AGNALFSSNPGSGSTAVNTFNO 975
 DB 937 ALAVQGIASSODDLTGILPLSDEFSAVITKQDTGWKVDIYKN--TNGLEKDDQLSE 993
 QY 976 N-----LSDAFESSGAKYTSDFWGTIOFKRPDELIONGFTSOVARFVNQSLNL 1028
 DB 994 NVKRRDNGLVPIYNEG-----IYDIMGKRVDFANSVLQARNLTDKVDVEYINNPDLQSF 1048
 QY 1029 VDETPANAGTNRVNVVDPDGNLTNQLPLKVOIQYLDGKRY--DAKLK-----NNNLVPS 1082
 DB 1049 EKFTAFDNRQAMLVGEKTSDT--LYAKPKREIYLDGNFYGDSKIAIPLINDPSRI 1105
 QY 1083 YNNGALPSWVVPATAGSTGLIATMIILGLAIGPLRAQRLODKGFTTEKKVDTLA 1142
 DB 1106 FAGFALPSPVIVPVSQSSVGLIILLIIGLIGIPMYVVRKODSEFVDVKKVDTLT 1165
 QY 1143 AVGSYVKRITITOTANYKKRPAALGAKSSDKRPAAKRPAAPKAPASSPAKRTGK 1202
 DB 1166 AVGSYVKRITITOTANYKKRPAALGAKSSDKRPAAKRPAAPKAPAPPAAPR----- 1210
 QY 1203 SGAPTKPTAPKPA 1215
 DB 1211 ---PVQP--PKKA 1218

RESULT 3
 ADP1_MYCA
 ID ADP1_MYCA STANDARD; PRT; 1122 AA.
 AC 049379; 049437; 053351;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
 GN MGCL.
 OS Mycoplasma gallisepticum.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxId=2096;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=56;
 RX MEDLINE=96201559; PubMed=8613358;
 RA Keeler C.L. Jr., Hnawo L.L., Whetzel P.L., Dohms J.E.;
 RT Cloning and characterization of a putative cytodhesin gene (mgcl)
 RL from Mycoplasma gallisepticum.
 RL Infect. Immun. 64:1541-1547(1996).
 RL (2)
 RP SEQUENCE OF 1-12 FROM N.A.
 RC STRAIN=56;
 RA Hnawo L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
 RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
 RL (3)

DN Adhesin_P1 precursor (Cytadhesin P1) (Attachment protein).
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID:2104;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=88057593; PubMed=3119495;
 RA Su C.-J., Tyson V.V., Baseman J.B.;
 RT "Cloning and sequence analysis of cytodhesin P1 gene from Mycoplasma
 pneumoniae";
 RL Infect. Immun. 55:3023-3029(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=88297153; PubMed=2841195;
 RX Imahime J.M., Denny T.P., Loeschel S., Schaper U., Huang C.H.,
 RA Bolt K.F., Hu P.C.;
 RT "Nucleotide sequence of the P1 attachment-protein gene of Mycoplasma
 pneumoniae";
 RL Gene 64:217-229(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himelereich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 RN [4]
 RP SEQUENCE OF 1301-1520 FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=88154763; PubMed=2450165;
 RA Dello S.F., Su C.-J., Horton J.R., Baseman J.B.;
 RT "Identification of P1 gene domain containing epitope(s) mediating
 Mycoplasma pneumoniae cytoadherence.";
 RL J. Exp. Med. 167:718-723(1988).
 RN [5]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=21088919; PubMed=11271496;
 RA Regula J.T., Ueberle B., Boguth G., Goerg A., Schnoelzer M.,
 RA Herrmann R., Fienk R.;
 RT "Towards a two-dimensional proteome map of Mycoplasma pneumoniae";
 RL Electrophoresis 21:3765-3780(2000).
 CC -1- FUNCTION: THE PROTEIN IS THE MAJOR ADHESIN MEDIATING THE
 CC ATTACHMENT OF THIS MYCOPLASMA TO RESPIRATORY EPITHELIUM.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M18639; AAA25424.1; -;
 DR EMBL; M21519; AAA88325.1; -;
 DR EMBL; AE000002; AA895661.1; -;
 DR EMBL; X07191; CAB37298.1; -;
 DR PIR; A41480; A41480.
 DR PIR; S03725; IUYMAP.
 DR InterPro; IPR004940; Adhesin_P1.
 DR Pfam; PF03257; Adhesin_P1; 1.
 KW Cyoadherence; Signal; Transmembrane; Complete proteome.
 FT SIGNAL 1 59
 FT CHAIN 60 1627 ADHESIN P1.
 FT TRANSMEM 1527 1547 POTENTIAL.
 FT SITE 1403 1415 CYTADHERENCE EPITOPE.
 SEQUENCE 1627 AA; 176270 MW; 14FA2CBA56E6116 CRC64;

Query Match	5.14; Score 328; DB 1; Length 1627;
Best Local Similarity	21.34; Pred. No. 1.9e-09;
Matches 268; Conservative 147; Mismatches 404; Indels 442; Gaps 66;	
QY	251 GNGOIT--NDPIA-----GTKTITDN-----QNPSTENGMPGANN 286
Db	506 GNGHVTSAHHAPLSLISGVFRVRYNATGSAATVGTMPALLFSGVNNQOTDGLKDPFNN 565
QY	287 RYDSOLNVKHKIKTSPOLDERINNTNSRIGNRNNNSKFVIG---VDNPAHVTFPTD-- 341
Db	566 RM-----FEVPRMAVAGAKFVGR---ELVLAGTITMDTAVPRILYDEL 608
QY	342 GTRKFNFTNQTGELVINDFEL-----DAPI-----LPKDLH--PDWYNL 377
Db	609 ESNLNLVAQGCGLLREDQLFTPPYGMANRPDLPIGAMSSSSSSSHNAPYTHHNPMDQDR 668
QY	378 YIQKKILPNDVNTAVVMPVGRVSGTNAD-----GMDFCGNGQITN-TDPIAQT 426
Db	669 PIQ-----NVYDAFIKPEW---DKNGDKDAKIYIPRYSGMW---AMQVYVNSNKLTDQ 716
QY	427 KTTTD--NONSTFNS-----GAMP-----GANN-----RYDSOLNVK--- 457
Db	717 PLSADFNENAYOPNSLFPAILNPBELLALPDKAYKGENEFANEXERFQKLTIVAPTO 776
QY	458 -----HRIKTSFO-----LDEKFVPEWTGSEENKNTRLATGSLPSNERWIL 501
Db	777 GTMNSHFSPTLSRSTGFLNGLVSDQVLDQVLPVIGNGRYG-----NNHR----- 822
QY	502 DTPGTPOVTLKEDSVNWFSLYILNSVNSLST-----GDSITYFGT 542
Db	823 ---GVDDITAAQTASGSSGISTISTGSRSLFPFNSIGVGLKANVOATLGSSOTMTIGG 879
QY	543 SELPSLWYSPPTRLSDLTALNOYKT-----DDLEASSTGDTGTTNGTTTADSSGSGT 597
Db	880 S-----PRTLDOANL-QLMTGAGMNDKASSQGSQSEBNTKFTLSATGMQOQSG 928
QY	598 AGTGNNTTNSQ---TVSNPTLNTYRSFGIDSKPTSAKIDETNNAD--PNVIEARIYAEY 652
Db	929 TSAQNPDSLKODNISKSGDSLTTODGNAIDQ-----EATNTNLPNLTPT---ADM 978
QY	653 RLGIQNEIPRTNAGN-----FINPTLIGGVFTSGSRVYLARAYNDQORTGMPORPLY 706
Db	979 ---PNAISFTNNAQRAQLFELGLLGSIT-----PVLNRRGSDSNKQATDQKSY 1027
QY	707 VEGYLGYOQRTFTGFMYGYTK-LIN-----NSPYVLDLSPRGV---TETNQ 748
Db	1028 T-DLHSDQRTKLNLRPAGEVNGLLNLPALVETFCNTRAGSGSMTSPGIFRIPEDNN 1085
QY	749 FRRTSLIYP-----VMGGLTFEGARSTSNPIYTAQGD----- 782
Db	1086 DSKATLLTPGLAMTPRODVGNLVSGTIVSFQGLMWLV-----TFPDEVPRAGYIGLQ 1138
QY	783 -----PRESRIFOSGSDTYEIGISGLFDCI-----RNNLWVGYKAS 822
Db	1139 LTGLDASDAIQRALIMAPRPAARSGSNV-NRLGRVESYMDLKGVMADQADSDQSGSTTT 1197
QY	823 SFLLNSRNPNGL---EMIAATYYL-RSQIGLARTSGL----- 856
Db	1198 ATRRALPEHPRALAEFOVSVEASAYKPRTSSGGOQTSTSSPYLHLVYKPKYTQSDKLDD 1257
QY	857 -----PMQ-----OPRGTHOIVTSVPGQSFSSIKIIRLIPFGNQLMYLFLFNENKNS 904
Db	1258 LKNLLDPQOVATKLRQSGFDTH---STOPQPO-SLKTTPPVFG-----TSSGMLS 1303
QY	905 SVYTLRLADSSNPDPASSFPTSLIDVNEI-----GVILPLLDNSFYTVNAGNVALPSSN 960
Db	1304 SVLS-----GGAGAGGSGSGSGQSGVDLSPEKVSQMLVGQLPST-----SDGNMS--STN 1351
QY	961 PGSGSGSTAVNTPMQ-----NLSDIAPEGSGSAKTTSPWGTIIOQRKPDXYLIONPFSQVA 1015
Db	1352 NLAQNT---NTGADYVGVGRLS---ESNAKAKNDDVDGIYRIPPLAEILLDGEQOTADTG 1403

```

OY 1016 RNFPT-----NOSEFLSLVD-FTTPAN-----AGTNYRV-----VYDP 1046
DB 1404 POSKESPDQIDNRLEFTHPYVDLFPVLMVYDQYIPLEFDIPASVMPKMKVRLKLSF 1463
OY 1047 DGNLTNOMLPKVOIQYIDKRYDAKLKN-----NLYVTS-----YANEGA 1088
DB 1464 DFN--EQSLGLRLEFFKPD---QDTOPNNVQVNPNGDFLPILITASSQCPOTLFSFVQ 1518
OY 1089 LPSWVPVPAIGSTIGLIMITLIGLIGLIPLRQRLQDCKGFKTKKVDLTFAVGSY 1148
DB 1519 WPDYVLPALTYPIVIVYSTLGLAIGIPMKRKKQALKGAFALSNOKVDVLTAVGSVF 1578
OY 1149 KRIITOTANVKKRPPALGAGSGDKKPAAPAKKPPAKSPKSPKPGKSGAPRT 1208
DB 1579 KEIINT-GISQAPKRL-----KQTSAAKPGAPPPVPPKPGAKPPVQF----PKK 1625
OY 1209 P 1209
DB 1626 P 1626

RESULT 5
ADP1_MYCSE STANDARD; PRT; 1444 AA.
ID ADP1_MYCSE
AC P20796; Q49286;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein) (MgPa).
GN MGPA OR MGPA OR MG191.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=89173298; PubMed=2925238;
RA Dailo S.F., Chavoya A., Su C.-J., Baseman J.B.;
RT "DNA and protein sequence homologies between the adhesins of
RL Mycoplasma genitalium and Mycoplasma pneumoniae.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=90060815; PubMed=2583522;
RA Imanin J.M., Loechele S., Collier A.M., Barile M.F., Hu P.-C.;
RT "Nucleotide sequence of the Mgpa (mgp) operon of Mycoplasma
RL genitalium and comparison to the p1 (mgp) operon of Mycoplasma
RN pneumoniae.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RN Fritschman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RX Fritschman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RN Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.W.,
RX Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lincier T.S.,
RN Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RL "The minimal gene complement of Mycoplasma genitalium.";
RN [4]
RP SEQUENCE OF 106-177 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=92051396; PubMed=1945886;
RA Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RN "A random sequencing approach for placing markers on the physical map
RT of Mycoplasma genitalium.";
RL Nucleic Acids Res. 19:6027-6031(1991).
RN [5]
RP SEQUENCE OF 468-588; 627-723; 980-1162 AND 1235-1289 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;

```

```

RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RN "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- FUNCTION: THE PROTEIN IS THE MAJOR ADHESIN MEDIATING THE
CC ATTACHMENT OF THIS MYCOPLASMA TO THE CLILATED EPITHELIUM.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M31431; AAA25420.1; -
CC EMBL: U39698; AAC71410.1; -
CC EMBL: X61522; CAA43734.1; -
CC EMBL: U02208; AAD12501.1; -
CC EMBL: U01694; AAB01007.1; -
CC EMBL: U01779; AAD10599.1; -
CC EMBL: U02159; AAD12441.1; -
CC PIR: A30588; A30588.
CC TIGR: MG191; -
DR InterPro: IPR004940; Adhesin_P1.
DR Pfam: PF03257; Adhesin_P1; 1.
KW Cytoadherence; Signal; Transmembrane; Complete proteome.
FT SIGNAL 1 1444
FT CHAIN 31 1444 ADHESIN P1.
FT TRANSMEM 1353 1373 POTENTIAL.
FT CONFLICT 1160 1161 SO -> FA (IN REF. 4).
SQ SEQUENCE 1444 AA; 159651 MW; DBIDFF7A90FFFB8A CRC64;

Query Match 4.4%; Score 285; DB 1; Length 1444;
Best local similarity 20.9%; Pred. No. 2.8e-07;
Matches 270; Conservative 159; Mismatches 477; Indels 386; Gaps 68;

OY 24 SASGFEKQSD-----KSDNTQOLVQNRATLDANSVYLAGIGNGSLFNTVLRDQDN 75
DB 418 NTSQGFYVGDTHDKKDFKKNSSPIALPPEAFANIGMVAIG-----NSV----- 464
OY 76 FITANGTIIKLDFTKPLVGLDSDCGGYKQKQVSDYTSRNFQDQRTAV-YAL- 133
DB 465 FIFGNGHATKMT-TNPL-----SIGVERIK-----YT---DNFSKSVTGMPIYVL 508
OY 134 ---LYNDEANVHLKRINTNSNR-----IGNRNNNSKFIYVG---VDNPAH 172
DB 509 FEGLINPQTN-GLDPLGLTKMFEYVPRMAVSGVKWGN-----OLVLAGLTLMGDTAT 562
OY 173 VIRETDDGDK--FNEFTQGEIVNDEFLDAPILPKLHPDWNYNIQRIKLPNDVNTAV 230
DB 563 VRLKTYDQLEKHLNVAGQGGILAREDDLIETPY-----GMAV-----RDIP----- 605
OY 231 VMPVGRVSGTNADGMDGCGGQTTNDPTAQTRTTNDQNPSTFNSGAMPAN---NR 287
DB 606 GAW-----LQDEM-----GSKFGPHVFLNN 625
OY 288 YDSQLNVKRIKTSFQDERINTNSRIGNRNNNSKFIYGVDPNPAHVIRETDDG---T 343
DB 626 PDIQDQVNN-----DTVALISYKYNIDKL-----KHVYPRYSGLAWQ 665
OY 344 KPNFTNQ-TOGEIVNDEFLDAPILPKLHPDWNYNIQRIKLPNDVNTAVPMPVGRVSG 402
DB 666 LFNWSNKLNTPTLSANFVENDSYAPRSLF-----AAILNEDLLGL----- 706
OY 403 TNADDCGFDGCGQIINTPTDIACTKTNDQNPSTFNSGAMPANRRDQSLQNVKRIKT 462
DB 707 ---SDKFFYKENEFEADREFNQLSLNPNP-----NTNMAKRLNVNQRTT 752
OY 463 SFQLD-----EKPV-YRPMWIGSEENKNITRLATGSLPSNERYWILDIPGTPVLTLEKDSVN 517

```

```

Db      753 GPNLDSFPDQFLDFLPWIGN--GKPFNSPSPSTSSASS-----STPLPFTFSNIING 802
Qy      518 VESRL--YNSVNSLSFIDSDIYIGTSLPSLM-----YSPPTRLSDLTALNOXYTD- 570
Db      803 VASMTQHLKKNTR-----WFLPNFSPDIMGAGRYOSANOKNIPPEQVPSYN 855
Qy      571 ---IEASSTDNGTNGTTTADT-----SSGSTAGTGNNTTSOTVSNPTLN--TYR 619
Db      856 STPFDPNSDNKVTBPGSSSKPTYPALPNSISPTSDMINALTFTK--NNPQRNOLLR 913
Qy      620 SFGIDSKPTSAKKIDETN-----WADPNVIEARI--YAEYRLGIQNEIPTNAGN 667
Db      914 SL-LGTIPVLIINKSGSDNQFNKDSQKDKETETNGNPGCEVN-GLYNALLHTYGF 971
Qy      668 FIRNTIG---GVGF-----TSTGSRVLRASYNQDQPTGN-----FQPELYVFGY 710
Db      972 FGTNTNSTDPKIGFADSSSSSSSTLVGSLMWTSDVGNLVINDTSFGFOLGWFITF 1031
Qy      711 LGYQOTRTGTFWVG--TYKLLNNSPYDVLDSPRVGETNQFRRTSLTYPYMGGLYFEGA- 768
Db      1032 TDFIRPTG--YLGITLSSLOPOTIIMADP-----WTSFK-----GSTLSDGTP 1075
Qy      769 -----RSFSNTPYIRAGDTPESRSIFQSGYSDNTYEXIQSVLGFQDGIKNNLWGVKAS 822
Db      1076 KSLMDPTALKSLPNSSTTYDTNPTLSPFQLYQPNKAYQTNTY----NKLIEPVAT 1131
Qy      823 SFLNSRPNPNLEMIATTYLRSQIGLARTSGLPNOQPGTTHOIVISPGDQFSIKN 882
Db      1132 SAAT---NMTSLKLTLTKTKAKLGKTAASQGNNGGVSQITNTTITGNTISEGK 1187
Qy      883 ITIIPGNOLWYFLFTENNKKSVYTLRLADSNPDASSFSFTSLIDNEIGVILPLD 942
Db      1188 EETSIQAEHLKKFEPSKONKS-----EIGI-----GD 1215
Qy      943 NSFYYTANAGNVALESNPGSGSTYAVTFNQNLSIDIEESGATYSDFMCTIOPKPD 1002
Db      1216 STFTKMD--GKLTGVSTP-----LVNLI-----GGATSDSD--TEKISFKG 1256
Qy      1003 EYLIONG--FTSGVARNFTVNSOFLNS-----LYDFTPA--NAGTNRYAVVDPGNTJN 1054
Db      1257 NQIDFNRLTTLPTVELFDPRNTMFVYDQYVPLVNLPSGDAQSIRLKI-----STSVEN 1311
Qy      1055 LPLKQIQYLDGK-----YDPAKILNNMLVTFSYNNFGALPSWVPVTAIGSTLGLI 1108
Db      1312 QFLGVLEFEKDPQTQGFIPVLNASTSGPTVFQPFNQMA---DYVLPITYVIVYIIS 1368
Qy      1109 IILGLAIGIPLRQKRLQDKGRTTKKVDTLTAANGSYKKITITOT--ANVKKKPAALG 1166
Db      1369 VTLGLTIGIPMHRNKKALQAGFDLSNKKYDVLTAKAVSVEKEIINTGISNAFK- 1423
Qy      1167 AGKSGDKRPAAKAPAPAKPSAPKASPAPK 1198
Db      1424 -----LKQATPTKPT-PK--TPPKP 1440

RESULT 6
OMPA.RICCN          STANDARD:          PRT: 2021 AA.
AC 052657; P95591; P95592; P95593; P95594; 052667; 052668; 052669;
AC 052670; 052674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (cell surface
DE antigen) (OmpA) (Omp A).
GN OMPA OR RC1273.
OS Rickettsia conorii.
OC Bacteria: Proteobacteria: Alphaproteobacteria: Rickettsiales:
OC Rickettsiaceae: Rickettsiae: Rickettsia.
OX NCBL_TaxID=781;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.:
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rOmpA.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.E., Roux V.;
RT Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rOmpA.";
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 001028; AAA17405.1; -
DR EMBL: AF008674; AAL03811.1; -
DR EMBL: U43794; AAB49549.1; -
DR EMBL: U43806; AAB49550.1; -
DR EMBL: U43806; AAB49551.1; -
DR EMBL: U45244; AAB49556.1; -
DR EMBL: U46918; AAB66653.1; -
DR EMBL: U83440; AAC35176.1; -
DR EMBL: U83443; AAC35179.1; -
DR EMBL: U83448; AAC35184.1; -
DR EMBL: U83453; AAC35189.1; -
DR InterPro: IPR006315; Autotransport.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRfam: TIGR01414; autotrans_bail; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1308 1308
FT N -> H (IN STRAIN MOROCCAN).
FT D -> A (IN STRAINS INDIAN TICK TYPHUS), M1
FT AND MOROCCAN).
FT N -> H (IN STRAIN MOROCCAN).

```

Query Match 3.68; Score 232.5; DB 1; Length 2021;
 Best Local Similarity 21.0%; Pred. No. 0.00023;
 Matches 278; Conservative 147; Mismatches 536; Indels 365; Gaps 64;

```

  QY 9 SYTLGGAVFGALGASAFSGFGQSKSDNTOLYONAKRTIDANSVRLAGLGONGSLFNTV 68
  DB 162 NTYGLGNALGSA-----NAALLIGSAA--PAKITLAG--NINGGGIIIV 202
  QY 69 LRDVDNFTTANGTITIKDSFTKPLVGLDSDCGYKVOIYSDVTSRNFQROTR 128
  DB 203 KTD-----AATNGTIGTNALATYVAGAGIA--FLEGALIKATTTKLTNNAVLTITNN 255
  QY 129 AYAALVNDKANV-HLKRIINTN-----SNRIQRN-----NSKFEVIGVDNPAHYI 174
  DB 256 AVLTGALDNTGVDNGLNLGALSQVYGNIGNTNALATISVGAGKATLGAVIKATTT 315
  QY 175 RFTDDGCKFENFTNQ--TGEIYNDFLDAPILPKDLHPDWXNLYIQRIKILPNVNTAVVP 232
  DB 316 KLTDNASAVTFNPPVVTGALDN-----TGANNNGIIVT 348
  QY 233 WP-----VGRVSGTNADGMEFGDNGQITNTDPIAOTKTT--TDNONPSTF--NSGAMGA 284
  DB 349 FTGDSIVTGNIGNTNA--LATISVGAGKATLGAAIKATTTKLTNNAVTFNPPVVTGA 407
  QY 285 NNRYDSQNLVHKRIKTSFOLDERINTNSNRIGNRN-----NSKFEVIGVDNPAHYI 337
  DB 408 ---IDNTGNANNGIIVT--FTGDSIVTGN--IGNTNALATISVGAGATLGAAIKATTTK 460
  QY 338 FTDDGTFENFTNQ--TGEIYNDFLDAPILPKDLHPDWXNLYIQRIKILPNVNTAVVP 395
  DB 461 LTNASAVTFNPPVVTGALDN-----TGANNNGIIVT 493
  QY 396 P-----VGRVSGTNADGMEFGDNGQITNTDPIAOTKTT--TDNONPSTF--NSGAMGA 447
  DB 494 TGDSTVYTNIGNTNA--LATISVGAGKATLGAAIKATTTKLTNNAVTFNPPVVTGA 551
  QY 448 NNRYDSQNLVHKRIKTSFOLDERINTNSNRIGNRN-----NSKFEVIGVDNPAHYI 507
  DB 552 ---IDNTGNANNGIIVT--FTGDSIVT--IGNTGNIN--ALATISV-----GAG 589

```

QY 508 QVTLKEDSVNVFSRLVNSLSFTGDSIYIFGHSLEPILMYXSPFRLSDLTALNOVK 567
 DB 590 KATLGAAIKATTTKLTNNAVTFN--NPVVTGALD-----TGNTNNSQVNSPTL 615
 QY 568 TDDIEASSTDNNGTTNGTTTADTSSGTYAG-----TGNTNNSQVNSPTL 615
 DB 630 ANNVIYFTNNSVTYGNIGNTNALATVNVAGAGIATLEGAVIKATTTKLTNNAVLTITNN 689
 QY 616 NYRSPGIDSKPTNSANKIDETNMADPNVIEARLYAEVRIQNEIPIITNAGFNITNTIG 675
 DB 690 NAVLTGALD-----NTTGVNDVGNLNLGALSQVYGNIGNTNALATISVGAG--KATLG 742
 QY 676 VGFSTGSRVYLAS-----YNGQRPNGF---QPLEY 706
 DB 743 AVIKATTTKLTNNAVTFNPPVVTGALDNTGNANNGIATFGDSTVYGNIGNTNALAT 802
 QY 707 VEGYLGOQTRTFGTFWYGYTKLLNSPBYDVLDSPRV-----GTE 745
 DB 803 VNVGAGLLRVGGVYKSNITNLTDNNAVTFNPPVVTGALDNTGNANNGIYFTGDSIV 862
 QY 746 TNGFRRTS--LTYPYMGGYLTEGARSFSTPYIRAQGDTPF--SRIFSGSYSDNTYXI 802
 DB 863 TGNIGNTNALATISVGAGKATLGAAIKATTTKLTNNAVTFNPPVVTGALDN----- 918
 QY 803 QSVLFGDGRNNLNVGVASSFLNSNRPNGLMIATTYLRSGIARTSGL--PNOQP 861
 DB 919 -----GNANNGIYFTGDSIVTGNIGNTNAL--ATVNVGAGVTLQAGGSLDANNID 967
 QY 862 FGTTHOVIVSPGDDFSSIKNIRITFPGNOQLYFL---FTENKK-----SSVYTLR 910
 DB 968 FGARSFLFENGPLDG-----GGNALIYPERKGAIVANGNNAIIVNTKLLTVAHLT 1016
 QY 911 LADSS--NPDASSF-----SPSLIDVNEIGVILPLDINSFEYVNAAG--NVALSSNP 961
 DB 1017 IGVVAELINGAGLFIIDNASAGVITLNMODIHFRALDVALVLSNLGVGNVNLILADL 1076
 QY 962 GSPGSAVAVTFENQNL-----SDIAFESSGAKTSDPWGTFQEPDEYLIONGFTSOVA 1015
 DB 1077 VAPVDGCVIVFDGAVGNLIGISNVA--GAARNIGDGN--KFNILLIYNVTTIDD 1130
 QY 1016 RNEVTNOSFL--NSLYDPTPA--NAGT-----NYRVVDPDGNLTNOMPLKVOIQVLD 1065
 DB 1131 VNLGQIIVLNNADTSTTAFNAGTIQINDATYTI--DANNGNINLTPAG--NIKF-- 1183
 QY 1066 GKYYDAKL-----KNNNLVTFSYNNGALPSWVVPFALGSLGILAIMTI-----LGIA 1114
 DB 1184 -AHADAQLIILQSSGNDRTITLGN-----IDPD--NDDEGIIVILNSVYAGKKLTIA 1232
 QY 1115 IGIPLRARKRLDDGFK-----TFPKVDTLFAVGS-----YKKIIT-- 1152
 DB 1233 GKTFGGAHKIOLDIVFKGSDGFGVAGTTFMTNIVLIDITGOLBLGATFANVVLFDVAVL 1292
 QY 1153 TQYANV 1158
 DB 1293 TQTGNI 1298

RESULT 7
 OMPB_RICPR STANDARD: PRT: 1643 AA.
 ID OMPB_RICPR
 AC 053020; Q9ZCM0;
 DT 30-MAY-2000 (Rel. 39, last sequence update)
 DT 30-MAY-2000 (Rel. 41, last annotation update)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DE (surface protein antigen B precursor (168 kDa surface-layer protein)
 DE (surface protein antigen B precursor (168 kDa surface-layer protein)
 DE (antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;

RN [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN-Breintl.
 RC MEDLINE-91045972; PubMed-2122457;
 RX Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RA "Characterization of the gene encoding the protective paracrystalline-
 RT surface-layer protein of Rickettsia prowazekii: presence of a
 RL truncated identical homolog in Rickettsia typhi.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Breintl.
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Madrid E;
 RX MEDLINE-99039499; PubMed-9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Neeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RL mitochondria.";
 RL Nature 396:133-140(1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN-Breintl.
 RX MEDLINE-92114896; PubMed-1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
 RL the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [5]
 RP CLEAVAGE SITE.
 RX MEDLINE-92104668; PubMed-1729180;
 RA Hachtschmidt T., Messer R., Cleplak W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RL membrane protein of rickettsiae: identification of an avirulent mutant
 RL deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M37647; AAA26390.1; ALT_INIT.
 DR EMBL: AF161079; AAD42234.1; -;
 DR EMBL: AJ235273; CA15140.1; -;
 DR PIR: D71630; D71630.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter.1.
 DR TrEMBL: TIGR01414; autotrans_Darl.2.
 KM Antigen: s-layer; Cell wall; 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1338 32 kDa BETA PEPTIDE.
 FT CHAIN 1339 1643 120 kDa SURFACE-EXPOSED PROTEIN.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINTL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINTL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINTL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).

FT CONFLICT 191 201 TTQAPLTGA -> INSRSSSYHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735DFE392E6346CC CRC64;
 Query Match 3.4%; Score 216.5; DB 1; Length 1643;
 Best Local Similarity 18.0%; Pred. No. 0.0012;
 Matches 231; Conservative 168; Mismatches 412; Indels 463; Gaps 59;
 14 GGLAV----FGALGASFGFKQS-----DKSNDNTQLVMOARTLDANSVRLAGONGSLF 65
 159 GGAAINANDLSGLSGITFAAPSVLEFNILNPTQ--EAPLILGANS-KIVN-GGNGTL- 213
 66 NTVLADVDNFTTANGITIKKDSFTKPLGLDSDCGGKVKQIVSYTTSRRNFDOR 125
 214 ----NITNGFIQVSDNTEFAGIKTIN-----IDDCG----- 240
 126 QTRAYVALLVNDEAVHLKRINTNSRIGNRNNSKFEVIGVDNPAHVIKFTDQ--TKF 183
 241 -----LWFNSTPDA-----ANTLNIQVGNNTIN-----FNGIDGKGLVYKNGKAATFE 285
 184 NPTNQTQGEIYVDFLDAPILPDLHPDWYNYLQKILPNDVNTAVVWMP--VGRVSG- 240
 286 NVTGLTGLNKKGITELNFAVAGKL-----ISQGAANNAVICTDNGAGRAAGF 333
 241 -TNADDG-----MPCCGNGQITNDPIAQTITTDN 270
 334 IYSVNDGNATISGVYAKNNVIOANAGQVTEHIVDVGIGGTTNFTADSKYITEN 393
 271 QNPSTFNSGAMPGANRRYDSOLNVKHR--IKTSFOLDEIRINTNSRIGNRNNSKFEVIG 328
 394 SNEGSTNENGL-----DQIVPEPTKILKNGFIDGVYNNNGTAGVITFNANGALVYS 446
 329 VD-----NPAHV--IRFTDGTKEFTNTQT-----QGEIYVD 358
 447 TDPNIAVTNINAIEAGAGVELSGIHIELRLNGSGIFKLADGVINGVQNALMNN 506
 359 FILDA-----PIPKRLHPDWYNYLQKILPNDVNTAVVWMPVGRVSGTNADGKF 410
 507 NLAAGSIQLDGSAITIGDIGNGVAAALQHTTLNDA-SKLALDGANIIGANVGAIH 565
 411 DCGNG--OITNPD-----PIAQRTTNDONPSFNSGAMPGANRRD 451
 566 PANGTKITLNTQNNIYVNFDDITDTKTVGDVDSLLNTQTLTNGSIGVYANTKIL 625
 452 SOLNVKHKIKTSFOLDEKFEVPEWMTGSEENKNTIRLATGSLPSNERWILDIPTGPQV 510
 626 AQLNI-----GSSK-----TILNAGDAVINE--LVIENNGSVQLNH 659
 511 ---LKEDSVNFSR---LYLNSVNSLSIGISIVYFGSELPMSLYFPPRLSLDTL 563
 660 NTYLTITKTIANOGOIIVAADPLNNTTLADGTNL-GSAENP-LSTIHFAKKAANDSI 717
 564 -----NOVKTDIEASTDNGTNTGTTTADPSSSGTG--AGT----- 600
 718 LNVGKGVNLIYANNITINDANVSLH--PESGTSIYSGIVGGQGGKLNLLINDCTYK 775
 601 --GNTT-NTSQTVSNFTL-----NTYSFSGIDSKPTSANKIDETNMAPDVNTEARIYAEVR 653
 776 FLGDTTFNGCFKIEKSLIQISNNYTDHVESADNGT-LEFVN-TDPIYTLNKGAV- 832
 654 LGIQNEIPLITNAGNFRNTIGVGVSTGSRVYLAASYGNDGAPPTGPFGLYFEGYLG 713
 833 FGVLVKVIITSGPNTIFENIGNV-----IVHGIAANS-----ISF 868
 714 OQPRGTFYVGYTKLLNPNPVLDSPRVGTETNGRRISLYPVMGYLTBERGARSFN 773
 869 ENASLCTSLF-----LPSGIFLDVLT----- 889
 774 TPYIRAGDTPESRSIFQSGYSDNTYEVYIOSVLGFDGIRNNNLNVGVKASSFLNSNPEN 833

DB 890 -----IKSTVG-NGVDNENAPLVVSGIDSM----- 915
 QY 834 GLEMIATYLRSGIGARTSGLPNOOPGTHQVTSVSPGOFSIKINRIIFPGNOLM 893
 DB 916 -----INNGOIIIGDKKRIITLISGDSNITVNAANTLYSG----- 949
 QY 894 YLEFTENNKSSVYTLKLDSSNDP-----ASSSFSP-----TSLIDVEIGVILP--LL 941
 DB 950 --IRFTKNNQCTV--TLSGGMPNNNGTITGLGNGSPKLVQVFTDYNNLSIIANNVT 1006
 QY 942 DNSFTY-----VNAGNVALFSSNPGSPSYNAVTFENONLSIDAF 982
 DB 1007 INDYTLTGTGGIAGTDFDAKITLGSVNGNANVRFVSTFSDPSMVAIQANK----- 1059
 QY 983 EESGAKYISDFWCTIQEFKPDDELYLQNGFTSQVARNFVYNOSFLN---SLVDFTPAAGTN 1039
 DB 1060 -----GIVTY-----LGNALVSNIGSLDTPVASVRFNGDSGAG 1093
 QY 1040 YRVVDDPQNLNLTNOLPLKVOIQYLDGKYDAKIKNNLY---TFSTYNN----- 1085
 DB 1094 LQ-----GNIVSON-----IDFGTYMLTINSVNIIGGTTAINGEIDLLTFNNLI 1138
 QY 1086 -FGALPSPVTPAIGSTLGI 1104
 DB 1139 FANGSTMGDNISITTLNV 1158

RESULT 8
 OMPB_RICUA STANDARD; PRT; 1656 AA.
 ID OMPB_RICUA
 AC 006653; 2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rompB)
 DE (romp B) [contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_Taxid=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein romp B of Rickettsia
 japonica";
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
 SIMILARITY).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AB003681; BAA20138.1;
 CC InterPro: IPR006315; Autotransport.
 CC InterPro: IPR005546; Autotransporter.
 CC Pfam: PF03797; Autotransporter; 1.
 CC TIGRPFAM: TIGR01414; autotrans_bar1; 2.

KW Antigen: S-layer; Cell wall. 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1338 32 kDa BETA PEPTIDE.
 FT CHAIN 1339 1656 POLY-GLY.
 FT DOMAIN 528 533
 SO SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;
 Query Match 3.48; Score 215; DB 1; Length 1656;
 Best Local Similarity 19.08; Pred No. 0.0014;
 Matches 285; Conservative 161; Mismatches 492; Indels 562; Gaps 74;
 14 GGLAV-----FGALGSASFGEKOS---DKSNDTQLVNQAARTLANSVRLAGLQNGCSLF 65
 160 GGAAIANDLSDGLTDFGAASTLVDFLANPTQ---KAPILADNALLYN-GANGTL- 214
 66 NTVLRVDNDFITRANGTIIKLDSFTKPLYGDLSDDCGCVKQVQVSDYTTSSRRRPPQR 125
 215 -----NVTNCFIQVSDKSPATYKAIMIGDQG----- 241
 126 QTRAYVALLVNDEANVHLKRIINTNSRIGRRNNSAFVIGVDNPAHVIRETDDG--TKF 183
 242 -----FMFTNA-----TNANALNLQAGCTIINNGTDGIRGLVLSKNGAATDF 286
 184 NFTNTOGEI-----VN-----DFT-----L 199
 287 NVTGSLGSLGKGIIEINTVAINQGLIANAGPANAIVGTNNGAGRAAGFVSVYDNGKATF 346
 200 DAPILPKDLHPDWNYLQIKILPNDVNTAVPVPVRSVGT-----NADD 245
 347 DGQVYAKD-----MVIOSANANGQVNEFHIV-DVG-IDGTTAFKTAISTVAITONSNE 397
 246 GMEFDGCGN--GQTTNTDPIAQTFTTDNONPSTFNSGAMPAN-----RYDSQLVNKKR 297
 398 GTTDFGNLAQVYTPDTHLTGNETGANNPAGVTTFANQGLTASASADANVAVTNN 457
 298 IKRSFOLD-----ERINTNSRIGRRNNS-----NSKRFVIGV----- 329
 458 I-TAIRASGVGVQVLSGTITAEIRLNGASVFKLADGVINGKYNQVTLGVGLAAGAIT 516
 330 -DNPAHY-----IRFTDGTG-----FNF-----TN 349
 517 LDGSATITGDIQNGGGGALLOSTITLANDATKTLTGAGNIISANGGTINFANGCTIKLT 576
 350 QTOGEIVNDF-----IIDAPILPKDLHPDWNYLQIKILPNDVNTAVPVPVRSVGT 399
 577 STONNIIVDDDLAIATDQGVVDSLSL-----TNAQTLTISGTIGI 617
 400 VSGTNADDGMEFDGCGNQTITNDPIAQTFTTDNONPSTFNSGAMPANRR--DS 452
 618 IGANNTILGQFNIGS-----SKTLNGNVAINELVYGNNGSVQFAHNTLYLTR 666
 453 QLVNKKRIKISFQIDKEFYVPFWTGSSEKNITRLATGSLPSNERYWILDIPGPVYLK 512
 667 TTNAAGDQKILF-----NPVYNNNTTLAAGT-----NIGSAAANPLAEIN 705
 513 EDSVAVFSLRLYNSVLSIFIGDSIYIFGT-----SELPJLMYSEFTRLSDTLNQV 566
 706 FGSKARADYLVN-----VGEGVNIYATITTTDANVGSFVFNAGKNIYSGIYGGQ- 757
 567 KTDIDIASSTNGTIT-----NGTIT-----TADTSSGSTAGTGNNTN 605
 758 QGNKEFTVALDGTGVKFLGNATFNGNTTIANSTQISGVYADFASADGIGIYEFVN 817
 606 TSGTYSNPLNTYRSFGIDSKPTSAKIDETNMADPNVIEARITAEIRLQIONEPITNA 665
 818 TGPI--NVTLN-----KQAVPVNALKQITVSGPGV-----VYNE--IGNA 854
 666 GNF-----IRNTI-----GCVGFSTGSRVYLR-ASYNGDQRPPTNGPFLYV 707
 855 GNVHGAMTDIAPENSSLAGVLELDSGIPENDAGNTIPLTIKSYGTNTAGCFSPSVIV 914
 708 FG-----YGYOQTRTG-----TFWYGYKLLNSPYDVLSPRY--- 742
 915 SGVDSVIADQVIGDQNNIVYGLGSDNGIIVNATVYLAGIGTINNNGCTVTLGSGVNT 974

```

OY 743 -----GTEYNFRRTSLT--YPMGGLTEEGARSESNTPYIRAGDTPESRIFQ 791
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 975 PCTVYGLGIGASAKFOVTTTDDNNLGNIA-----TWTT--NGVYTTTGIA 1025
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 792 SEYSDNTYVYIOSVLGFDGIRNNLNVYKASSFLNSRPNGLMEIAAT-----TYL 844
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1026 GGIAGTDFD---GKITLGSVNGNANVRFDAGIFSNST-----SMIVTTKANNGVTY 1075
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 845 -----RSQIGLARTSLPNOQ-----PFGTHOYI---SVSPGOFSS 879
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1076 GNAFVGNIGSDPTPAVSVPFTGSNNAGLKGNIYSGVIDEGYINLGVNSNVLGSGTFA 1135
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 880 IK-----NIRTFPGNOLWYFLFTNNKSSVYLRAD-----SSNDPSSPS 924
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1136 INKIDDLTNTLTFAGGTSTW-----GNNTSLEITTLTANGNIGIIVABEQAVATTT 1189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 925 PSLIDVNEIGVILPLLDNSFYTVNAGNVAFSSNPGSPGYTAVN---TFNQLSDIA 981
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1190 GTTITWOD-----NANAN---FSGTQ---TYTLTGGAARFNGTLGSPN 1227
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 982 FEGSGAKYTSDFWGTIOFKPDEYLT-QNGFTSOVANRFTYNSF-----LNSLVDF 1031
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1228 FTVTGSNREFVNY-GLIRANODYVITRTNNAENITVNDITNSPFGGAPGVGVNVTFFVNA 1286
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1032 TPANAGTNY-----RVVVDPDGNTLNONLPLKQVIOYLQDKKYDAKLKNNN 1077
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1287 TMTAAVNNLLAKNSADSNFVGTIYTDISAITNAQOLDVADIQOLG----- 1335
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1078 LVTFSYNFGALPSWVVPFAISGTGLIAMIILG-----LAIGI---PLRAQRLODK 1128
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1336 -----NRLGALRYLCTPEMVGSEAGAIIPAAVAAGDEAVDNVAYGIMAPPEYIDAHQSK 1389
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1129 GKRTTFKK-----VDLTF-----AAGSVYKKITQTPANVKKRAALGAGSGDK 1173
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1390 GGLAGYKAKTGTGIVIGLDTLANNMIGAIG-----TKKT-DIKHODY-----KKGDK 1437
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 9

OMP_RICCN STANDARD; PRT; 1655 AA.

```

AC 09KKK3; 09KK98; 09XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN OMPB OR RC1085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fourrier P.-E., Barde V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN 121
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer membrane protein ompb (compb).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN 131
RP SEQUENCE OF 353-1655 FROM N.A.

```

```

RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent Rickettsia of the spotted fever group.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: AE008659; AAL03623.1; -
DR EMBL: AF123721; AAF34124.1; -
DR EMBL: AF123726; AAF34129.1; -
DR EMBL: AF149110; AAD39533.1; -
DR PIR: E97835; E97835.
DR InterPro: IPR006315; Autotransport.
DR InterPro: IPR005546; Autotransporter.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRPFAMs: TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1335 1635 32 kDa BETA PEPTIDE.
FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 K D -> L (IN REF. 3).
FT CONFLICT 776 776 F -> S (IN REF. 3).
FT CONFLICT 1159 1159 E -> D (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

```

Query Match

3.28; Score 204; DB 1; Length 1655;

Best Local Similarity 19.58; Pred. No. 0.0054; Matches 272; Conservative 172; Mismatches 484; Indels 468; Gaps 71;

```

OY 15 GLAVFGALGASGFQKQSDKSNBDTQLVNQARTLDANSVRLAGLQNGSLFNTVLRDVD 74
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 272 GRVLISLKNAAADPFWTGSIGNLKGIIEFVAVANGOLKANAGANAIVIG-----N 325
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 75 NFPTAANGTIKID-----SFKPLVGLDL-----SDDCGYKKQIVSDYTSRRNRPDR 125
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 326 NCAGRAAGFVAVSDNCKVATIDGOVYAKDWIOSANAVGVNRRHIVDGTDTTFAFKTA 385
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 126 QTRAYVALLVN-----DEANVHLKRI--NT-----NSNRIGNRNNSKFIYGVNDPAHYI 174
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 386 ASKV--AITONSFGTTDGNLAAQIIVFTMTLNGFTGDASNP-----NTAGVI 435
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 175 RFTDGTGKFNFTNOTGEIYVNDPILPAPILPKDLHPDWNYLXIQRIKLPNDVTVAVPMP 234
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 436 TFPANGLTASASADAVAVYNNI-----TALASG 465
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 235 VG--RVSGTNADD-----GMFDGNGQITNTDPIAQKTKTTTNDONSPSTFNSGAMPGAN 285
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 466 AGVVALSGTHAAELRLGNAGSVFKLADGTVIN-----GKYNQTFALVGALAGT 514
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```


Accession	Source	Standard	PRT	AA
Ys89_CAEEL	Ys89_CAEEL	STANDARD	PRT	3178 AA
009624	009625	0969D4		
01-NOV-1995	(Rel. 32, Created)			
28-FEB-2003	(Rel. 41, Last sequence update)			
15-SEP-2003	(Rel. 42, Last annotation update)			
De	Hypothetical protein ZK945.9 in chromosome II.			
ZK945.9/ZK945.10				
GN				
Caenorhabditis elegans				
Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditoidea;				
Rhabditidae; Peloderinae; Caenorhabditis.				
NCBI_TaxID=6239;				
OX				
RN				
SEQUENCE FROM N.A.				
RC	STRAIN-Bristol N2;			
WLKun	Wilson-Sproat J.;			
RL	Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.			
RA	Durbin R.;			
RP	Revisions.			
CC	(2)			
CC	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
CC	1- SUPRACELLULAR LOCATION: Integral membrane protein (Potential).			
CC	1- SIMILARITY: Contains 1 GPS domain.			
CC	1- SIMILARITY: Contains 1 PLAT domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-slb.ch/announce/			
CC	or send an email to license@isb-slb.ch).			
CC	-----			
DR	EMBL; Z48544; CAB70192.1; JOINED.			
DR	EMBL; Z48582; CAB70201.1; JOINED.			
DR	EMBL; Z48544; CAB70201.1; JOINED.			
DR	WormPep; ZK945.9; CE25697.			
DR	InterPro: IPR002111; Cat_channel_TrpL.			
DR	InterPro: IPR005821; Ion_trans.			
DR	InterPro: IPR001024; Lipoxigenase_LH2.			
DR	InterPro: IPR003915; PKD_2.			
DR	InterPro: IPR000203; PKD_cys_rich.			
DR	InterPro: PF01825; GPS; 1.			
DR	Pfam: PF00520; Ion_trans; 1.			
DR	Pfam: PF01477; PLAT; 1.			
DR	PRINTS: PM001433; POLYCYSTIN2.			
DR	SMART: SM00303; GPS; 1.			
DR	SMART: SM00308; LH2; 1.			
DR	PROSITE: PS50095; PLAT; 1.			
KW	Hypothetical protein; Transmembrane.			
KW	Hypothetical protein; Transmembrane.			
FT	TRANSMEM 13 30			
FT	TRANSMEM 51 73			
FT	TRANSMEM 2139 2161			
FT	TRANSMEM 2348 2367			
FT	TRANSMEM 2390 2412			
FT	TRANSMEM 2451 2468			
FT	TRANSMEM 2483 2505			
FT	TRANSMEM 2567 2589			
FT	TRANSMEM 2836 2858			
FT	TRANSMEM 2839 2961			
FT	TRANSMEM 2976 2998			
FT	TRANSMEM 3038 3060			
FT	TRANSMEM 266 1196			
FT	DOMAIN 1105 1241			
FT	DOMAIN 2071 2120			
FT	DOMAIN 2182 2305			
FT	PLAT			

SQ SEQUENCE 3178 AA; 344726 MW; F8293436D03666CD CRC64;
 Query Match 3.1%; Score 202; DB 1; Length 3178;
 Best Local Similarity 19.9%; Pred. No. 0.016; Indels 266; Gaps 41;
 Matches 201; Conservative 129; Mismatches 414;

```

OY 254 QITNDPIAQTKTITDNONP-----SFGNSGAMPANRRYSQLVNFKIKTSFOL 304
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 532 QQSSTTSETTSSDGTNPDEYFEKATTFYDSR--SVNLINSGGLI---IGVQSI 586
OY 305 DERITNSNRIRNRNNSKFV-----IGVDNPAHVIRFDDGKFNFTNOGGEIVN 357
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 587 ECTSPSSNMYSTTKDAGCFKTSVSMPLRGTT--YVGPFG---NYT--FRATMT 638
OY 358 DEILDAPILPKDLHPDWNYLIQR-----KILPNVNTAVYVWPVG 398
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 639 D-----DKVYTYTANVYIQEYSSSTIESESSSAVASSTSPSPSSSTLSTVY 690
OY 399 RVSGTNADDMEDCGNGQIT-----NTDPIAQTKTITDNONPSTFNSGAM 443
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 691 EESSRSDSTTSSAGSTTLOESTTSESTDSSTTISDPTSSSSPSTLADS--- 747
OY 444 PANNRYDSQLNFKIKTSFOLDEKFYPEWTGSENNITRLATGSLPSNR---YVI 500
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 748 -----TSTLSVD--QDFILDSGLSWNETRHEDSINIVPLPNAITPTERSQTFEC 797
OY 501 LDIPGTVTLKEDSVNVSRLVNSVNSLSFIDSIYFGSELPSLWYSPPTSLDL 560
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 798 RVNSREPLIKESTCLANSYVLANATYSSNIPIQIEF---LVGISTEFRIMTDL 853
OY 561 TALNOVKTDIDIASSTDNCTTNGTTTADTSSGSTAGCTGNTTTSQTVSNPTLTYS 620
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 854 TTM-QVSHIFTLNVADSTSEVSTSTGSSSSSAISTGSIESTST----- 903
OY 621 FQIDSKPTSAKIDETNMADPNVIEARIRAGIONPIPTNAGNFRNTIGVGFPS 680
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 904 --LEASTDASODSSSTSDS-----GTTSDSTTIDSSNPSSTSDSSGLSQ 948
OY 681 TGSRVVLASVNGDORPTGNFQFLVFGYLGQRTGTFWYGTAKLNNSPYDLNDSP 740
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 949 TPS---DSSASDSKRTTTPD-----DATERPYDVLN 981
OY 741 RVGETTNOFRRTSLTYPVMGVLT---EGARSFSNTPYRAQDTP---ESRSIF 790
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 982 LTMNEFTVYYSNP-----YITPIPNKEPALITAMCOCORNDSSQFVLKESNCLT 1034
OY 791 QSG-----VS-----DNTRYIGSVLGFDDGRNNLNVYKASSPLNSRPNP 832
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1035 EFGKNGAISASVFNPMTSFVPATGTYEFLIN-----TNRASESASHIFTMNVLP 1087
OY 833 NGLMIATATYLRSGIARTSGLPNOQPFQTT---HQVYSVPGDQFSSIKNIRTFPG 889
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1088 TTTTTPPTTVSSSDAGKGTGTGTGTGTGSGSASATLTSTGDAVAST----- 1138
OY 890 NOLWTFLEFNENKSSVYTLRLADSSNPASSFSFTSLIDVNEIGVILLPLDNGTYT-- 947
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1139 -----TSGSGSGSSTGSGAGSGGTASGSGSGS--SGTGSIDV-----NSKRTTA 1183
OY 948 VNAAGNVALLFSSNPGS---PGSYTAVNTFNON---LSQIAREGSAKTTSDWGTIQK 1000
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1184 LMGDTGSGTATTPGSHLDGGSTSGSGSDSSGSSVSTKSSGSGSTSGSSDSSG-- 1239
OY 1001 PDEYLQNGFTSOVARNFTNOSFLNSLVDFTPANAGTYRVVAVDPD--GNLTQONPLKY 1059
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1240 -----NGAFSATQPSRTTKTRSSLATVSPISAAE--QALIIDQKADVMNQ----- 1284
OY 1060 QIQYLDGKYDAKLKNNLVTPS--YNNFGALPS---WVPPAIGSTGLIAMIILGL 1113
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1285 ---LAG-IMDSASNNLSINTSSLSLQISSLPAADLVYVAOSLISNTLTKIPGVNMSV 1339
OY 1114 AIGIPRAQRKLODGKFTTFKVVDTLTAAGSVYKIIITOTANVKKRA 1163
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1340 DV-----LKTLDONIAITNSELADENAKVITKLANVMTSA 1375
  
```

RESULT 11
 MRP_STRSU
 ID MRP_STRSU STANDARD; PRT; 1256 AA.
 AC P32653;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Muramidase-released protein precursor (136 kDa surface protein).
 GN MRP.
 OS Streptococcus suis.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NC NCBL_TaxID=1307;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=serotype 2 / D282;
 RX MEDLINE=92267650; PubMed=1587602;
 RA Smith H.E., Vecht U., Gielkens A.L., Smits M.A.;
 RT Cloning and nucleotide sequence of the gene encoding the
 RT 136-kilodalton surface protein (muramidase-released protein) of
 CC Streptococcus suis type 2.
 CC Infect. Immun. 60:2361-2367(1992).
 CC RL -i- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X64450; CAA45781.1; .
 CC DR PIR: A43829; A43829.
 CC DR InterPro: IPR005877; Gpos_YsIRK.
 CC DR InterPro: IPR001899; Gram_pos_anchor.
 CC DR InterPro: IPR006192; LPXTG.
 CC DR Pfam: PF00746; Gram_pos_anchor; 1.
 CC DR Pfam: PF04650; YsIRK_signal; 1.
 CC DR TIGRfams: TIGR01167; LPXTG_anchor; 1.
 CC DR TIGRfams: TIGR01168; YsIRK_signal; 1.
 CC DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
 CC KW Cell wall; Peptidoglycan-anchor; Repeat; signal.
 CC FT SIGNAL 1 47
 CC FT CHAIN 48 1226
 CC FT PROPEP 1227 1256
 CC FT REPEAT 663 1256
 CC FT REPEAT 663 681
 CC FT REPEAT 839 861
 CC FT REPEAT 953 1006
 CC FT REPEAT 1064 1064
 CC FT REPEAT 1089 1142
 CC FT REPEAT 1143 1195
 CC FT SITE 1223 1227
 CC FT MOD_RES 1226 1226
 CC FT SEQUENCE 1256 AA; 135795 MW; DCF765242F14341 CRC64;
 Query Match 3.1%; Score 199; DB 1; Length 1256;
 Best Local Similarity 19.4%; Pred. No. 0.0068;
 Matches 266; Conservative 148; Mismatches 522; Indels 436; Gaps 67;

```

OY 60 ONGSLNFVLRVDVDFITRANQTIKLSFTKPLGLDSDCGGKVKQIYSDYTSR 119
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 80 ENTSVAATSTEVE-----KAKAVLDQVISESLGL-----GQELAKTEDATLAK 127
OY 120 NRFD-ORQRAYAYALLVDEANVHLKRIINTNSRIGNR--NNNSKFEVIGV----- 167
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 128 AIEDAQTKLAARAILADSEATVEQVEAOVAAKVANEMALGNELQKTYVDGLTALDHY 187
OY 168 --DNPAHVIRFND--DQTKFNFTNQGELVNDPILDAPILPKDL-----HFDW-- 212
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
  
```

```

Db 188 APDTASTLKVGDEGLDSTTATPNAEP--NGAIAPIHPLTIDODIKATSEPNWT 245
OY 213 ---NLYIQRK-----ILPDVN--TAVPEPVRGVRSGTNADGMFDC 250
Db 246 PESIDLYSNKNMNASSTYKGAEDVAYIRYSLDDSSSTAVLAELVSRITGDVLEKTYLER 305
OY 251 GNGQITNDPIAOKTKITTDNON-----PSTFNSGAMPGA-----NNRYDSOLANKHRK 299
Db 306 GE-SVTFESH--TKVNAANSNITVTYDTSLASANTGALKFSANDVSTIIIVPAYQIN 361
OY 300 TSEFOLDERINTSNRIGNRNNNSKRVIGVDNPAHVIRFDDGCKNEFTNOTQGEIYNDF 359
Db 362 TT-----RVTESGKV--LATYGLQTIAGQVYTPSSVRVFGYDVATTTAAVQGPYPRKG 415
OY 360 IIDAPILPKDLHPDMVNLXIOKILIPVNTAVY-----PWPVGRVSGTNADGMFDCG 413
Db 416 VYLAGTVOKDT---VOYKVIIEIEND--QAVLKFYLDPTYKGEVDMRG-----460
OY 414 NGQITNDPIAOKTKITTDNONPSTFNSGAMPGANKRYSOLNVKHRITKTSFOLD-----467
Db 461 ----TDTGTIELLT-----SPTTYKVGTI-----YD--YVNSKITAPFTIDPTKNM 504
OY 468 -----EKF--VYPEWTESEENKI-----TRIATGSLPSENERWILDIPGTQ 508
Db 505 VFKESEONEQSKRYVIAOMSGDETCKIGIKIYATQVWTKLGTNMGWF-----556
OY 509 VTLKEDSVNFSRLYNSVNSISFGDSIYITGTSLSPLWYYSFPTSLDLTALNOYKT 568
Db 557 -DYDDQAGI-----KFNKNGP-----WPAGVONLIRNMTPTAVET 592
OY 569 DDIKASSIDNCT-----TNG--TTTADTSSGSGTAGTGNTN---TSQTSNPLN 616
Db 593 TYIKESSKYDVIYEYIDDGKQIVASVDPFKSALGTEYNDVRRPASIVADGVY 652
OY 617 TYRSGIDSKPTSAKIDETNMADPNVTEARITAEVYLGIONELPITNAGNIRNTIGCV 676
Db 653 FKKEYKSDSAKTTGVAVAGTTVK-----YTEKAGSVN-----V 687
OY 677 GRTSGSRVY--LRASYNDORPTGNFOPFLYVGYGLOOTRGTFWYGYKLLNNSPD 735
Db 688 NFEDINGKVIKAPVEDENDAKPGVNYDTL-----DOKLASTFEGKEKILVPAGDP 740
OY 736 VLDSPRVGETNOFRITSLITVVMGYLEGASFSWTPYIRAGDTPSRISFOSGIS 795
Db 741 V--GKVG-----KNNLIEVG---NMT---AKGIDPTGKIEAGVKN 774
OY 796 DNTYEXI---QSVLGEFGDIRNLL-----NVGKASSFLMSNRPN-----831
Db 775 EVIYVYRAVGSVYVNYKDEGNVICKDPETDVSADPVDDATYTTDKKNEIITKDSRYV 834
OY 832 -----PNG--LEMIATITLRSOIG--LARTSGLPNOQ-----PFGTT--HOVI 869
Db 835 LVPSKTDGEENGKVIETITVTYVYOKVANNMPEIPNVETDRPVPYFPFTEPERID 894
OY 870 SVSPGDQFSSIKRITIFPGNOLMYLFTNENKSSVYTLRLADSSNDPASSFSPTSLI 929
Db 895 PTTGQTN-GEVNPINPY-PG-----YT-----PYDPDNTPIAKPI 927
OY 930 DVNEIGVILPLDINSFTVNAAGNVALFSSNPG--SPGSAVAV-----NTFNONISDLA- 981
Db 928 DENDPG-----KGVYPTPTPNPGVDPIPVYPAKVVYTNHVDEGNPIAP 972
OY 982 -----FEGSgAKYTSDFWGTIOKPEDEKILONGCFISOVARNFVNO--SFLNS 1027
Db 973 QEGEGRPKNSIGYEFTGKTVTDGEGNTLHYKTPPEVKG--TYVAVYTTEDGVIKE 1029
OY 1028 LVDETPAN--AGTNYRVVVDPOGNTLNONLPLKV-----QIOYLDGKYDAKLKNNLY 1079
Db 1030 PYVDPTSPGEPYD-----TTDKKPTITFKGEYELVAVDGTENGKYESEETVY 1080
OY 1080 TFSYNNFGALPSVNVPTAIGSLGLAIIILGLAIGIPLAOKR-----LQDGKFTFP 1134
Db 1081 TYVYRKVEPAPKVVYTNHVDDE-----GNPVAPQDEGIRPKNSIGYEFTG 1126

```

```

OY 1135 KKVDTLTAAGSVYKRIITQITANVKKRPAALGACKSGDKKPAAPAKP-SAPKAS 1193
Db 1127 KVTYDEGNTTHYK-----TPAKKVYTNHVDEGN-----PIAQEGGTTPKRO 1172
OY 1194 SPA-----KPGKSGAPTKPTAPKPAKPIAP 1222
Db 1173 ISGEYVYTVVDEGNTTHYKLSKNKPTTPKERTPAKQAGKTAGKAOIP 1224

RESULT 12
P3P_LACTIC STANDARD; PRT; 1902 AA.
ID P3P_LACTIC
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P11I-type proteinase precursor (EC 3.4.21.96) (lactocoeplin) (cell wall-associated serine proteinase).
GN PRP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN-SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos F., Simons G., Slezew R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
  cell envelope-located serine proteinase."
  J. Biol. Chem. 264:13579-13585(1989).
RL [1]
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
  GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
  specificity, although some substrate preference have been noted,
  e.g. large hydrophobic residues in the p1 and p4 positions, and
  pro in the p2 position. Best known for its action on caseins,
  although it has been shown to hydrolyze hemoglobin and oxidized
  insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (potential).
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib.ch).
CC -----
CC EMBL: J04962; AAC03533.1; ALT_SEQ.
CC HSSP: P00782; 2SPT.
CC MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPXTG.
DR InterPro: IPR003137; PA.
DR InterPro: IPR002029; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR Pfam: PF02225; PA.1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR TIGRfam: TIGR01167; LPXTG_anchor.1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING.1.
DR PROSITE: PS00137; SUBTILASE_ASPI.1.
DR PROSITE: PS00136; SUBTILASE_HIS.1.
DR PROSITE: PS00138; SUBTILASE_SER.1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT P11I-TYPE PROTEINASE.

```

QY	Db	1192	DLVADKHHDTLLKFFDTKVRLLMAFTFIDKLFENNSDDTSEATIKVTGTVSADTPTVNGH	1251
QY	Db	646	--ARVAEYRLGIONEIPITWAGNFIRNTIGCVGFTSGSKHVLRAVNSGDRPRTGNQF	703
QY	Db	1252	TVVALDQNHHSV--DVPVNGDGTIKVTATDKNGNTTEBEKTTTSSVDRPM-----	1301
QY	Db	704	FLVYGLYVQOQRTGTFTWGTGYKLLNSP--YDVLSPRVG--TETNQFRRTSLYPRVW	759
QY	Db	1302	-----LKSVPFDGCVKFGGTNKFNAATSAKFYD---PKGIATIGTKVHKHPPTTLIWD	1350
QY	Db	760	GGVLTREGARFSNT-----PYIRAGSDPTRESIRPSGSDMTYELIQSVLGGDG	810
QY	Db	1351	GKQIPRIKDDLFSTFLDLGLTGOKPFGVGVGDPTQNT-PQ-----	1390
QY	Db	811	IRNNLVNGKVASFSFLNSRNPNPNCLEMIATTVYLRSGIARTSGLPNQPFCT--THQ	867
QY	Db	1391	-----EALSPFI-----LDVAAPTLLSDSSDTDAVYINNDPFOITGTATDMAQ	1432
QY	Db	868	VISV----SPGDQFSSIKNRITFFPGNOLMFLFTENNK-----SSYTLRLDSSNPD	918
QY	Db	1433	YLSLSINGSSVASQYEDI--NINSGKPGH---MAIDQPVKLEGGKNVLTPAVVDSBENT	1486
QY	Db	919	ASSSEF-----SPTSLIDVNEIGVILPLLDNSPYTYNAD--GNVALFSSNPGSPG	965
QY	Db	1487	TTKNTIYYEPEKKTLLAAPTVPYPTTEPAQIVTL-----TANAAATGETVOYSADGC--	1537
QY	Db	966	SYTAVNTFNONLSDIAFEGSGAKYTSDFWGTIOKPD-----EYLIONGFTSQVA	1015
QY	Db	1538	-----KTYQDV--PAAGVITITAN--GTFKRSKDLVGNESPAVDYVTNLIKADPPA	1584
QY	Db	1016	RNEVYTNOSFLNSLVDFTPAANAGTV-----RVVYDPDG---NLTONLP	1056
QY	Db	1585	OLOAKQELNLLIASAKTSLASGYKDDATTTALAAATOKAQTALDQTNASVDSLTPGARNR	1644
QY	Db	1057	LKVQIOYLDGKY--YDAKLKNNNLV-----TSSYNNFGLPSWYV	1094
QY	Db	1645	LQTAIINOLAALPDAKRTSLINLOQSVKDALGTDLGNOTDPTSTKTFI---AALDDLVA	1700
QY	Db	1095	PTAIG-----STLGILAMILGLAIGIPLRORKLDDKGFRTTPEK-----VDPLT	1141
QY	Db	1701	QAQAGTQTDQDLQATLAKIIDEVLAKLAEGIKAAKTPAEVGNAKMAAGKTMVADLDTLT	1760
QY	Db	1142	AANGSVYKKIITGTANWKKRPALGAGSGDKKRPAAAKPAPAKPS-----	1188
QY	Db	1761	SGQAS-----ADASDKLAHQALQSLTKTKVAALVEAKTVGCGDTGTSDKGGQ	1811
QY	Db	1189	---APKXSSPAKPTG-----PKSGA--PTKP-----TAPKRAAPK	1218
QY	Db	1812	GTPAPADGDTGDKDGDEGSGPSSGGINPTKPAATTTSTTDDPTTRNGOLNTHSGKALPK	1869

RESULT 13

OMPA_RICRI STANDARD: PRT; 2249 AA.

AC P15921;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmp A).

GN OMPA.

OS Rickettsia rickettsii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=783;

AN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=R;

RA MEDLINE=90354033; PubMed=2117568;

RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;

RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences."

OY	357	NOFLIDAPILPEKDLHPDWNYLIQRIKLLPNDVNTAVPMPVGVYST--NADDGMD-CGN	414
Db	1136	-----PVY-----	VTGADTNGANNIGYTFNGN 1159
OY	415	---GGITNDPIAQTKTITTDNONPSTFENGSGAMGNRRYSOLJNKHRIKTSFOLDEKE	470
Db	1160	SVTGGIGIGITNMLA-----TVNNGACITLIQADGSGLAANN-----IDFGARSTLEFNGPID---	1209
OY	471	VYEWEMGSEENKITERLATGSLPSNERKMYLLDIPGEPVYLKEDSVVFSRLYVNSL	530
Db	1210	-----GGKRAIPYEFKGI-ANGNALLINV--NTKLLASHLITITVAEINIGAGNLF	1259
OY	531	SF---IGDSIYIFGISELPSLMTYSPRR-----LSDLTAL---NOYKTDLEASSTONG	579
Db	1260	TIDASVD-VTILNAONI-----NFRADSVLYSLNLTGVGVNNILLADIVAPGADG	1312
OY	580	TTT-NGTTTADTSSG-----STGAGTGNNTT-----	SOVSNPPLNTY 618
Db	1313	TVFENGGVNGLNAGSNVAGTARNIGGGGNKKNFTLLIYNAVTTDDVNLBSEIONVLINKN	1372
OY	619	RSFGIDSKPYSAN---KIDETWADPNVIEARIEAYEYRLGIONEIPITNAGNFRMTIG	674
Db	1373	ADF---TSSAFNNGAIQIDAIY-----TIDA-----	NNGN-LINTPAG 1407
OY	675	CYGFSTGSRVVLKASYNQDQRP--GNFOPFLYFGL-----GYOQRTG--FPWY	723
Db	1408	NIOFHADAQOLVLNSSGNDRITLIGANIDPDMDGSIYILNSVTAGKLTITAGKRTG-	1466
OY	724	GYKLLNSPYDVIDSPVGTETNQPRTSLTFPVNGYLTREGARS-----FSNTPYIR	778
Db	1467	GAHKIQTILFKGAGCSTAGT---TENTNIVLDING--QLEIGATTAVVLEFNAVOLT	1521
OY	779	AGGDPESHESIPQSGVSDNTEYIQSVLGEQDGIKRNLLNVCVAKSSPLNSRNPNGLEMI	838
Db	1522	QIGNT-----GGFLD-----FNAKKGNVTLNNVNN---AGAVQNTGGIN-NGTLLV	1564
OY	839	AATTYLRSGIARISGLPNOQPGTTHQVTSVSPGQFSIKNRTIPFGNOLW---YF	895
Db	1565	LGASNLRNRYNGIAMLK-----VAGNVTLAKGKVKIGIQTGT---NTLLPAHF	1613
OY	896	LFENENKSSVYTLRLADSSNPDASSFSPTSLDVNEIGVLLPILDNSF--YVNAGNV	954
Db	1614	NLIGSINKTGQALKL-NFMNGSVSGVVGTA---ANSVGITTAGATSPRASSVNAKGT	1669
OY	955	AL-----FSSNPGSPGSYT---AAVTFNONLSDIAEFGSGAKYTSDFWGTIOEKPEYL	1005
Db	1670	TLGGTTSFATFTNIGAVTLAKGSITSFKNVTATSFVANS--TINFSNLAENSN--I	1725
OY	1006	IONGFTSQVARNFTY--NOSFLNSLV-----	1029
Db	1726	TGGGTITLIGANQVITYTGTSFDTLTLNTTFDGAAGSGNILLKSGSTDLDSQVRL	1785
OY	1030	-----DFTPAANGTVNRVY--VDPOGNL---TNONLPLKVOIOYLDGKYIDAKLKNNL	1078
Db	1786	VYATNFDPMNNSIPDTKTYIASFETAGLKPISKEVKTITIN-----NDNRF	1832
OY	1079	VTFSYNNGCALPSWVVPATGSLGT---LAIMITLG-LAIGIFLR-----	1120
Db	1833	VDFITD-----ASTLTLRAEDIAADVLDGDFAPROGFLANIPNANIKKSLE	1878
OY	1121	-----AQRIQDQGFRTTEKKYDVTLLTAAVGSYKKIITQTA-----NVKKRP	1162
Db	1879	LMEDAPNGSDARQAFNNGLMTPIQEDADATHLIDQVFKPSDTIAAVNQVVASINSSTN	1938
OY	1163	AILGAG---KSGDKRPPAAAKPPAAPAK	1186
Db	1939	TALNAAMDVQSGNKGPVSSGDEMDAK	1966

01-OCT-1993 (Rel. 27, Created)
01-OCT-1993 (Rel. 27, last sequence update)
01-NOV-1997 (Rel. 35, last annotation update)
MSB2 protein (Multicopy suppression of a budding defect 2).
MSB2 OR YGR014W.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
defect.";
RL Yeast 8:315-323(1992).
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -1- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC -1- SIMILARITY: SOME, TO YEAST HKR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: M77354; AAA34798.1; -;
DR EMBL: Z72799; CAA96997.1; -;
DR PIR: S25370; S25370.
DR SGD: S0003246; MSB2.
DR GO: GO:0005887; C: integral to plasma membrane; IMP.
DR GO: GO:0005034; F: osmosensor activity; IMP.
DR GO: GO:0000283; P: establishment of cell polarity (sensu Sacch. .; IGI.
DR GO: GO:0006970; P: response to osmotic stress; IMP.
KM Transmembrane; glycoprotein; Repeat.
FT DOMAIN 698 816 7 x 17 AA TANDEM REPEATS.
FT REPEAT 698 714 1.
FT REPEAT 715 731 2.
FT REPEAT 732 748 3.
FT REPEAT 749 765 4.
FT REPEAT 766 782 5.
FT REPEAT 783 799 6.
FT REPEAT 800 816 7.
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;
Query Match 3.1%; Score 197; DB 1; Length 1306;
Best Local Similarity 18.5%; Pred. No. 0.0091;
Matches 234; Conservative 176; Mismatches 456; Indels 396; Gaps 56;
DY 12 LIGLAVFGLGAS-FCF-----KSDKSNDTQLVNOARTIDANSVRLA----- 56
DB 8 LLSLTVISGLARASPDFFFGNGTQOQSOSESOGVSFTNEA-SQDSSTSLVPAYSQ 66
DY 57 GL--GONGSLFNTVLRVDNFTTAANGTIIKIDSFTKPLXGLDLSDDGGRVYKQVSD 114
DB 67 GVSHSGATVTSATISSLPSTWDASS-----TSQTSVYASQESD---YAVNQ--NS 114
DY 115 YTSRNFQDQRTAVAYALLVNDEANYHLKINTNSNRIGNRNNSKFIYGVDPNPAHI 174
DB 115 WSASTNQLPSTSTSYAPFPFSADPAASVNAAS-----DVSTASVPIDTSANSI 166
DY 175 RFTDDGKRFNTQDGEIYNDPILAPILPKDLHPWYLYIQKILPNDVNTAVVWP 234
DB 167 PFT---TTSNIETTSAPLSD-----TP 187

235 VGRVSGTNADGKFDGNGOITNTDPIAOKTTTNDON--PSTFNGAMPGANRVDQ 292
DB 188 LISTMSADANVFSSAN-----PISALTTTDSSEFSDOTSTGALPVGSS----- 234
DY 293 NVKRIKTSFOLDERINTNSNRIGNRNNSKFIYGVDPNPAHIRETDGKRFNTNQ 352
DB 235 -----ADPSSSEILVQSSA-----DFSSPS 256
DY 353 GEIYNDPILDAPILPDLHPDWYNYLIQKILPNDVNTAVVWPVGR--YSGINADGAF 410
DB 257 PTTTDSLAAPIQTS-----SSFTTSAALPVSTDDGSSAPVPS 301
DY 411 DCGNGOITNDPIAOKTTTNDONPS-TFN--SGAPGAN-NRVDQLVNRKIKTSFQL 466
DB 302 MSAAGQIAS-----STDNTKSETSLTETVDGSSVSTVALAPLQITSTN 354
DY 467 DEKFVYP-EMTGEENKNTRLATGSLPSNERVWILDPGPQVTLKEDSVNFRILY 524
DB 355 SFSIVPSVSVFVPSQSSSDVASSSTANVSS--FSDIP--PQSTSGSVSV-----A 404
DY 525 NVNNSLSFICDSTIYICGTSPLWYYSFTRLSDLTALNQVKTDDIEASTONG--TJT 582
DB 405 QASALAF-QSTREYVGA-----ASTKSLSLSTT 434
DY 583 NGTTTADTSS-GSTAGTGNTN--TSQTVSNPTL-----NTYRSFGIDSKPT-----S 629
DB 435 SLOSTLIDSSSLASSSSDLDYGVSSASTIPLISASQASTSSFSVSVSVFVS 494
DY 630 ANKIDETNMADPNVIEARIYAEVRLIQNEIPTNAGNFRNTIGCVFTSGSRVLA 689
DB 495 QSSSDVASTAPSV-----VSSFSYSLQAGSSMTNPSSSTIYVS 536
DY 690 SYNGDQRPICNPFPPLVFCYLGQOTRTGTFWYGYKLLNPNPYDVLDSFRGTETNQ 749
DB 537 SSTGSEESAATAATLSG-----SSSTVYAGN--LQSQP-----PSTSSLSES 580
DY 750 RRTSLIYVVMGVLTERGARGFNT--PYIAQGDPEESRIFGSGSDNT--YETIOSVLG 807
DB 581 QATS-ISAUL-----ASSSVSTSPYTTAGASATBASLISTSAETSOVSYSOS--- 629
DY 808 FDGIRNNLVNKAASFLNSRPNPGLMEIAATYLRSGIQLARTSGLPNOQFCTTHO 867
DB 630 -----TVALQTSFPASS--TTGSE-----TSSQGFSTSV 659
DY 868 VISV--SPGDPSIKNIRTIFFGNOLMYLFTNENKSSVTLR-----LADSSNPASS 921
DB 660 LVQMPSSISSEFSPSQTTQM-----NSASSSSQVYITISGTILSOVSDTSVS 707
DY 922 SFSPTSLIDNEIGVILPLLDNSFFYVNAAGNALFSPNGSGSYTAANTFNQNLSDIA 981
DB 708 TTSSSSVQVSDTPV-----SYTTSSS-----VQVSDTPVSTTSSSVQVSD-- 753
DY 982 FEESGAKYTSDFWGTIOFK--PDEYLQNGFTSQVARNFYTNOSFLNSL--VDFTPANAG 1037
DB 754 ---TPVSYTSSSSVQVSDTPVSYTTSSSSVQVSDTPVSPSSRSVQVSDTPVPST 810
DY 1038 TNRVYVDPDGNLTNQNLPLKVOIQYLDGKYIAKLKNNNLVFPSTNFEALSWVPTA 1097
DB 811 SSRSSVQVSSSLQ-----PTTSSQ-----RFTJSTHGLS----- 842
DY 1098 IGSTLGLAIIILGLAIGIPLRAORLQDKGKFTFKVDTLAAAGSVYKKILIJOTAN 1157
DB 843 -----ESSVSQASSETTSSINATASETHSIQTAA 874
DY 1158 VKKKPALGAGKGDKKP--AAAKAPAAPK-----PSAPKASSPAKPTGKSGAPK 1208
DB 875 QSTTLSTFDANSSASAPLEVARTPTPSSKASLLTLPSTSSLSQVATWTNVQTSLTTE 934
DY 1209 PT 1210
DB 935 ST 936

FT	CARBOHYD	1128	1128	N-LINKED (POTENTIAL).
FT	CARBOHYD	1140	1140	N-LINKED (POTENTIAL).
FT	CARBOHYD	1146	1146	N-LINKED (POTENTIAL).
FT	CARBOHYD	1211	1211	N-LINKED (POTENTIAL).
FT	CARBOHYD	1300 AA:	132801 MW; E09E52C3F6472430 CRC64;	
SQ	SEQUENCE			

Query Match	3.0%;	Score 192;	DB 1;	Length 1300;
Best Local Similarity	19.7%;	Pred. NO. 0.016;		
Matches 265;	Conservative 147;	Mismatches 450;	Indels 480;	Gaps 74.

OY	2	NISKRLK-----SYLLIG-----	GLAVGALGSAPFGKSDSKNDTQLVNQA	45
Db	50	NLAQIKVPAITLTFNFGDASNGCRAGVTTPDANGTL-----ESASADAVATANNI		104
OY	46	RTLDANS-----VRLAG-----LGONGSLFNTVLADVDDNFITRANGTIIKLDSFKPLY	95	
Db	105	TAIRSAGAYVALSGTHAAELRIAGNSGIFK-----	LADGVINGKVNTALV	152
OY	96	GILDSDDGGKYKQIVSYDTTSRRNFDRORRAYALLVNDEN---	VHLKRINTSN	151
Db	153	G-----GALAAGTIILDS-----	ARTTGIDGAGAMAALOMITLAND	190
OY	152	-----RIGNRNNSKFVIYG-----VDNPAH---VIRPTDGRKFNFTNOGEIYNDF	--	197
Db	191	AKKTLEIGAN-----IIGAGGTTIDLQAGGTIKLT-----STQNUIYVFDEDLA	235	
OY	198	-----IIDAPIILPKDLHPWNYLIQRIKLPENDVNAVVPWRGVSQTINADGMPCD	250	
Db	236	IATDOTGYVDASSL-----	TNQOTTLINKRGITIGANKKITLGQRNI	276
OY	251	NGOITNPDIPIAOCTTTTDNONPS-----FNSGAMPGANRRY--DSOLNVKHRIKTSFO	303	
Db	277	GS-----SKTVLSNGVAIALNELVIGNDCAVOFAHDYTLIRTNAGAOKRIIF-	324	
OY	304	LDERINTSNRIGNRNNNSKFVIYGVDNPAHVIRFTDDGRKFPMTNOTOGELYNDFILDA	363	
Db	325	-NPVVNNGTLLAGTN-----LGSATNPBLAEINGSKGVND- TYLVNGEGV-----	369	
OY	364	PILFRDLHPWNYLIQRIKLPENDVNAVVPWRGVSQTINADGMPCDNGOITNPDI	423	
Db	370	-----NLV-ATNITTITDANVGSEFVNAG-----GTINVGSGVGOOGKNKFTVAL	413	
OY	424	AQTKTTDNQSPSTFNSGAMPKANRYKSQIAVVKHRITSFQIDREKPYEWEMTSBENKN	483	
Db	414	-ENGTYVEAFGNATFNPGNTTIAAN-----	STIQIGANTV-ACCVASAQGTG	457
OY	484	ITRLA-TGSLPSNEREYMLDPGPPOVTLKEDSVNFSLYANSV-----NLSLFIGDISIV	538	
Db	458	IYEENVTEGIYYT-----LNQAAPVNALKOITVSPGVAVINEIGMAGNHGAVTDTI-	511	
OY	539	IFGSELPLSMY-----SPTLRSDLTALNOVKRTDIEA-----	573	
Db	512	AFENSISLIGAVVFLPRGCIIPFDNAGNMPLIKSIYVKNKTAKGFEPVSVVLGVDYIADQ	571	
OY	574	-----SSTDNGTTTNGTTTAADTSSGSTAGT-----GNTNTSOFVSAPTNT	617	
Db	572	VIGDONNVIGLGISDNGIIVATVILVAGISTLINNOGITVLYSGVPNRPGV----	624	
OY	618	YRSFGIDSKTYSANKIDEETNMADP-----NYLEARIVAEXRLGIONEIPTLANGNFRMT	672	
Db	625	----YGLDT-GIGASKRKOVYTTTTDINNIGIAT-----NATINDVITYTTCG-----	668	
OY	673	IGGVETSGTSRVYLRASYNGDOPRPNQPPLYVFGVIGQQTRTQTFYCYTKLLNS	732	
Db	669	IAGICFODK-----ITLGSVNGN-----GNVR-----PADGIS-----NSTSMIGETLK-ANNG	711	
OY	733	PYDVLDSPRVGETETNORRRTSLTRYPVMGGLYTEGARBSFSNTPY--IRAQDIPESHISIF	790	
Db	712	TVTYLGNFVN-----IGD-----SDTPVASVRFETD-----SDS	741	
OY	791	GGSYGNFYKEYIQSVLCEDGIRNNLVNKAASSFLNSNRPNGCNLEMATATYTLRSQIGL	856	
Db	742	GAGLOGNITY--SQYIDF-----GTYNIGI-----VASN-----IILGGTVAINGKIDL	783	


```
QY 851 ARTSGLPNOQPGTTHQVIVSVPDQFSSIKNIRITIFPGNOLWYFELFTNENKSSVYTLR 910
Db 784 -----VTNLTLEFASGTSW-----GNNTSIETTLR 808
QY 911 LADSSNPDASSSEPTSLIDVNEIGVITLPLDNSFTYVNAAGNAL-----FSSNPGSPG 965
Db 809 LANGN-----IGHIV-ILEGAQVNTTTGTTTIKVODNANANFSGTQ 849
QY 966 SYTAVN---TFNONLSDIAFEGSGAKYTSDEFMGTIOFKPDEYLI-----ONGFTSOVAR 1016
Db 850 TITLIGGARFNGTLCSPNFAYTGSNREFVNY-SLIRANODIYITRTNNANENVYTIIDIAN 908
QY 1017 N-----FV--TNOSFLNSLYDFTPANAGTNY--RVVYDPDGNLTNQNPLK 1058
Db 909 SPFGAPGVDOQVNTTFVNATNTAAAYNNLLAKNSANSANFVGAIYVDTSAITNVQDLA 968
QY 1059 VQIQYLDGKYIDAKLKNNNLVTFSYNNFGALPSWVP-TA--IGSTLGIATKTIIG--- 1112
Db 969 KDIQAQLG-----NRLGALRYLGTPTETAEMAGPEAGAISAAYVAGDEA 1011
QY 1113 ---LAIGI---PLRAQRKLODKGFKTTFK-----VDTLTF-----AAYGSVYKKI 1151
Db 1012 IDNVAYGIWAKPEYPTDAHQSKRGLAGYAKTTCVVIGLDTLANDNLMIGAATG----- 1065
QY 1152 ITQTANVKKRPAALGAGKSGDK 1173
Db 1066 ITFT-DIKHODY-----KKGDK 1081
```

Search completed: October 10, 2003, 16:33:55
Job time : 34 secs

THIS PAGE BLANK (USPTO)

QY 362 DAPILPKDLHPDMYNIYORKILPNDVNTAVWPVGRVSGTNDADGMPGCGQITNTD 421
 133 DVPNNSMNOVNLKLLLEOPLTKRYTLNNS-----LAKKQKQRE--VHLSG----- 179
 422 PIAGKTTTNDONSTENSGAMPGANNRD--SOLNNKRIKTSFQDEKVFYPE----- 474
 180 ---QANMOWTSORNDLNNNSPNAASGFKLTGNAARKLSESMPTPEIDGTGKQKGD 236
 475 ---WTGSEEN--KNITRLATGSLPSN-----BRVWLLIPGTPQVTLKEDSV 516
 237 SSCMSSTEENEAKKNAPSVSGGSSSGTGNKYLTGKALBSIGLFFDQPR----- 288
 517 NVPSRLYNSVNSLSFGDSIYIFGTSELPMLWTSFPTRLSDLTALNOKTDDIAPST 576
 289 NVITOLYASTSKLATNHNHIVMGNSFLPSKMWV-----VERSAQ 330
 577 DNETTTNGTTTADTSSG-----STGAGTGNTNT-----SQVSNPLNTYRS 620
 331 EN--ASNKPFWANTMLDMGEDKQKQFVENQGLGKETTSTNSHNFHSKSTQP--ALIL 385
 621 FGID-----SKPTSANKIDETNMADNPTEARLYAEYRIGI 656
 386 SGIDSVNDQILFSGFKAGSVGYDSSSSSSSSSTKDOALAMSTTSLDSK--TGKDLV 443
 657 ONEIPITNAGNFINTIGGVTSTGSRVYLKASVNGDQPTGNFQPELYVFGIYGOOT 716
 444 TNDTGLNGPIN-----GFSIDQTFEFV--PYSGNHTNNGTGTPKITAPVKKDOKS 494
 717 RGTGFWGYTK--LLNNSPYDLDSFRVGT-----ETNQRRTSLTYPY--MGQY 762
 495 -----TVKINSILNATPLNSYGDEGIVGFDALGNVFNKNOERLBSRIDQIFVVG 546
 763 LREEGARFSNTPYIRAGDTPESRSIFQSGSDNTYELYSVGLGFDGILNNLNVGKAS 822
 547 VSNPNELR--AKSSADSTGSDTKVMSNQSRYLVPYNSGEGIIDAGGFRPNKASVT 605
 823 SFLNSRPNPNCLEMIATYILRS-----QIGLARTSGI--PNOQFEG--THQVLSVSPD 875
 606 TF-----SGLKSTAPDGFANSTANFVGL--KAGIDPPVSGKKANYGAVVLRGG 655
 876 QFSSIKNIRTF--PGNOLWYFLFTNBN-----KS 904
 656 V-----VRLNFPQND--SLSTDDNNIAPISFSPPTAESAVALDTTFKEVYVQES 707
 905 SVYTLRLADSNPDASSSESPT-----SLIDVNEIGVPLPLDINSFYTVAAGN--VALF 957
 708 GLMSYIFDSSILKPSHDOKPTPTDNGFVSITVSRGIELN--QDQATTLIDVAPASALAVQ 766
 958 SSNPGSPGSYTAV-----NTFNONTSDI-----AFE-----GS 985
 767 SGIOSTGOTLIGVPLSEFSFSAVLAKSDQNKIDIKNNNGLEFIDTOLINSVATNNGL 826
 986 GAKYTS--DEWGTIOKRPDEYLLQNGFTSOVARNFVTNQSFLNSLVDTPPANAAGTNRV 1042
 827 ASSTYENRVDAMGKVEFDNSVLOARNLVKYDEILINPELINSFRTFPAFEDOKATL 886
 1043 VYDPPGNLTNOLPLKVOIYLDGKYYDAKIKNNLVTESYN-----NPGALPSMV 1093
 887 VATKQ--SDTSLASPRIGFLDGFEDL--NSTIGVPLNIGFSPSRVPAGFALPAMV 940
 1094 VPTAIGSLGLIATMILIGLAGIPLRQARKLOKGFTEFKKVDLTJLAVGSVYKTIIT 1153
 941 IYVSVSSVGIILFILLVGLGIGIPMYKVKLQDASFNWPKKYVDLTJLAVGSVYKTIIT 1000
 1154 QTNVKKRPAALGAGSGDKRPPAAAKPAAPAKPSAPKASSPAPCTGPKGAPLPTAPK 1213
 1001 QTGKVKKAPBALKAANPSVKKPAALFKP--PVQP-----PSKPBGGQKAVYKSEETK 1051

RESULT 2
 J50069
 hypothetical p1 operon 130k protein precursor - Mycoplasma pneumoniae (strain ATCC 29342)

A: Variety: ATCC 29342
 C: Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 07-Dec-1999
 C: Accession: J50069; 573338
 R: Name: J. M.; Loecheil, S.; Hu, P.-C.
 Gene 73, 175-183, 1988
 A: Title: Analysis of the nucleotide sequence of the p1 operon of Mycoplasma pneumonia
 A: Reference number: J50069; MUID: 89211947; PMID: 2468577
 A: Accession: J50069
 A: Molecule type: DNA
 A: Residues: 1-1218 (1MA)
 A: Cross-references: GB:M20916; NID:g150138; PIDN:AAA88326.1; PID:g1196658
 R: Himmelreich, R.; Hilbert, H.; Ploges, H.; Pirkl, E.; Li, B.-C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A: Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumonia
 A: Reference number: 573327; MUID: 97105885; PMID: 8948633
 A: Accession: 573338
 A: Molecule type: DNA
 A: Status: nucleic acid sequence not shown; translation not shown
 A: Residues: 1-1218 (HIM)
 A: Cross-references: EMBL:AE000002; GB:U00089; NID:g1673651; PIDN:AA895660.1; PID:g167
 A: Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C: Genetics:

Query Match 8.7%; Score 556.5; DB 2; Length 1218;

Best Local Similarity 22.1%; Pred. No. 1.5e-21;

Matches 321; Conservative 176; Mismatches 483; Indels 473; Gaps 65;

QY 1 MNISKLKSYTLIGLAVGALGASAFGKSDKSDNTOLVNOARFLDANSVRLAGLG 60
 1 MKSKLKIRKLPLPLPGTSLA-----WTYLLDHNTLTPYT----- 40
 DB 61 NSLFTVLDVNDNFITANGTI--IKDSFTKPYGLDLSDDGCVKQVSDYTTSR 119
 41 -----PFTPLNGLDVVRRAHLHPSVELDKRVGDTKLVALVRS--ALVR 85
 120 NRFDQRTAVYLLVNDVNDVNHKIRNTNSRIGRNNNSKFYIGGVNDPAHYTRFTDD 179
 86 VKFDQRTS--SDQS-----NTNOMNLSDTQESKALNG-----SQS 120
 QY 180 GTKFNFTNTOGEIYVNDPILDAPIPKDLHPDMYNIYORKILPNDVNTAVWPVGRV 239
 121 G-----SSDTGSGNSQDPAHYVLFKKAAPRATV--VFEKRI-----KLALPY--VKQES 165
 240 GTNADGMEFDGNGQITNTDPIAQTKT--TDN-----QNPSTNSGAMGANNR 287
 166 QCSGQSGSGKGLKTLQDLLEQDPYTPYPNGLARVNGVADIVHFGSGGESSWNSQ 225
 QY 288 YDSQNLNKH--RIKTSFOLDERINTNSRIGRNNNSKFYIGGVNDPAHYTRFTDDGT 343
 226 -RSQKGLKNNPKPAVATGKLDK-----GRAYKRLNESMPVVEPLDS----- 266
 344 KFNFTNOTGEIYVNDPILDAPIPKDLHPDMYNIYORKILPNDVNTAVWPVGRV 402
 267 ---TEGKG-----KD--ESSWKN-----SEKTAENDAPLGVYVGS 298
 DB 403 TNADGMEFDGNG-----QITNTDPIAQTKTITDNONPSTFN----- 439
 299 GAACSASSILOGNSNSGLKSLRSAPVVPSPSSNQTLSISNAPVGPQAVVSOAGG 358
 QY 440 SGAPGANNRYDSOLNKHRIKTSFQDEKVF-----YPMWTSSEENKNITRLATGSLPSN 495
 359 ATAASVNRRTASDTRFSKYLNTAOLHOGVYVPELEKMGNN-----GIGVYASR 410
 QY 496 ERYWLLDP-----GTPQ-----VTL-----KDSV--NVS 520
 DB 411 QDATSTNLPAAAGSQTGLGTGSPREPALATTSQARAVTVAGPLRAGNSSTFDALPNYIT 470

QY	521	RLVANSVNSLSFJDSIDSLYICOTSELLPSLWY-----	SFPRRLSDLT	561		
Db	471	QLYHTSTPAQALAYLNGQIVYMGSDBVPSLMWVVGEDQESGATWMAKTELWGDOKOF		530		
QY	562	ALNOVK--TDDLEASTNGTGTTCGTTTADTSSG-----	STGA--GTGNTNT	606		
Db	531	VENOLGKRBDSNSDKSNSNLKAQGLTQPAYLIAGLDVAADHLVAAPKAGAVGIDMTTDS		590		
QY	607	SQIVSNPLTYRSFGIDSKPTSAKIDETWADPNVIEARIYAEVRLGIONEIPITNAG		666		
Db	591	SASTYNOALAMSTTACGLDSD-----	GGYKALVENTAGLNP	627		
QY	667	N-----	FIRNTJGVGFSTGSRVYLARASVNGDQRPQPLFYGVGLGYQOT	716		
Db	628	NGELTLLDTRFAYVYIPVSGMKGGSONNEVOYTTYPVKSQKATAKIA-----		673		
QY	717	RTGTFMWGYTKLLNSP-----	YDV-----	IDSPRVGETNOFRRTSLYP	757	
Db	674	-----	SLINASPLNSYGDDGVYFEDALGINFKNLEERLPSRTDQL----	LYV	718	
QY	758	VMGFIYLE-----	EGARFSN-----	TPYIRAGDIPES-----	RS	788
Db	719	-----	GIVSELSKASRENAQSTSDSNSNTKYKWTNTASHYLPVYYYS--	ANPEAGNRRA	774	
QY	789	IFQSGYDNTYE-----	YIOSVLGFDGIRNLNVYKASSFLSNSRPNGLMEIATY		842	
Db	775	EQRNGVATISTESQATDGFANSLNFG--	TGLAGVDPAVANGHKNTYSAVILVRG--		829	
QY	843	YLRASQIGLARTSGLPNQ-----	-----	PEGTTHOYIVSYSPGQFSSIKNR	884	
Db	830	-----	GVRLNFPDPTDKLSDTKNSEPIISFYTPPGSAESAVDL-----	TLTKOV--	876	
QY	885	TIFPGNOLWYLFYN-----	ENKKS--SVYPL--RLADSNSPDASS-----	SFSPT	926	
Db	877	TYAESGLMEFTFPNGEKPTYDGAQQQYKNNKGAVITVSTGTGIEFNDDATYTLISQAP			936	
QY	927	SLIDVNEIGV-----	ILPLDNSFTYVNA-----	AGNVALESSNPSGSGSYTAVNFNQ	975	
Db	937	ALAVONGIASODDLTGILPLSDDFSVAITKDOTWTGKVDTYKN--	TNGLFEEKDOLSE		993	
QY	976	N-----	LSDIAFEGSGAKYTSPPFWGJOEKPRDEYLQONFTSOVANFPTNOSFLWSL		1028	
Db	994	NVKRRDNGLVEITYEG-----	YDIMGVDFPAANSVLOARNLDTYDEVINPDILOSF		1048	
QY	1029	VDFTPANAGTYRVVDPDGNLTJNQNLPLKVOIYLDGKXY--	DAKLE-----	NNNLVYFS	1082	
Db	1049	EKFPAPFQDNRAMLVEKXTSDYT--	LTVKPKIEYLDQENFYGEDSKINGICPLINDPERSI		1109	
QY	1083	YNNNGCALPBMWVYPALIGSTLGLAIMITLGLAIGLPLRAQKKLDDGKTKTPFKKYDTLTA			1142	
Db	1106	FAGPAALPSPWVIPVSGSVGILITLILGLIGIGIPMKYVKRLDDSSFDVDEKFKYDTLTT			1165	
QY	1143	AVGSVYKKIITQYANVKKKPAALAGSGDKKPAALAAKPAAPAKSPAKSSPAKPTGPK			1200	
Db	1166	AVGSVYKKIITQTSVVIKKAPSA-----	KAANNAPAKPVPVPAATATARP-----		1210	
QY	1203	SGAPTKPPAKPA	1215			
Db	1211	---EVPQ--PKKA	1218			

RESULT 3
T18346
MGCl protein precursor - Mycoplasma gallisepticum
C:Species: Mycoplasma gallisepticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18346
R:Keeler Jr., C.L.; Hnatow, L.L.; Whetzel, P.L.; Dohms, J.E.
I:Infect. Immun. 64, 1541-1547, 1996
A:Title: Cloning and characterization of a putative cytidinesin gene (mgcl) from Mycoplasma
A:Reference number: Z18861; MUID:96201559; PMID:8613358
A:Accession: T18346
A:Status: Preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA
A:Residues: 1-1122 <KEE>
A:Cross-preferences: EMBL.U34842; NID:g1022701; PID:g1022703; PIDN:AAB02987.1.
C:Genetics:
A:Gene: mgc1
A:Genetic code: SGC3

Query match	7.8%;	Score 501.5;	DB 2;	Length 1122;
-------------	-------	--------------	-------	--------------

Matches 311; Conservative 174; Mismatches 470; Indels 423; Gaps 68;

QY 5 KKLKSYTLIGLAVFGALGSASFGEFKQSDKSDNTQLVNQARTLDANSVRLAGLGQNGSL 6

Db 2 KKL-IFKLSVGITPLALIGLSFGLAIVSGAKPNNLKPVNQVGEMNSQ-----GQS-NL 5

QY 65 ENTVLRDVEDNF-----ITANGTIKLDSTKPLYG-----LDLS I

```

Db      53 LEKARRWRNSNFTSLSIDGTNPGALVLTGSKSISRID-----LYGNVIWTFDPGNTNDLT 1
          |  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

0Y 101 DDCGY-----KVKQIVSDY---TTSRNRFDOROTRAYALLVNDKANVHLKRI 1

D_B 108 GKGVEYDANKLTAFSGDVSENVSDI.SSKTVVEATODPEDPNVEYLLMPDAA----- 1

0Y 147 NTNSNRTGNRNNSKFKVTGGVDNPAHVTRETDGSKKFNFTNOTGETV-NDFTLPAPIIP 2

Pb 161 -----VVOEOKTKDOVEFNYSMDAP--- 1
 ::: | | |
 ::: | | |

306 KRI HROVNYI V IOPKTI BNDVNTAVI/BNDVVC BYSCSTNADOC-----MEDCGN 3

```

183 -----| : | | :: | | : : | |
      A C D T E A E C A M P B A C C C C C C A A C C C A V A B A A S C T A B I Y E E C N ?

```

[illegible][illegible][illegible][illegible]

1

— — — — —

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

DD 304 0N0F VMEF MWQ I I INEGB DI RNDGMY I DE 00NG I MYNNNEBNEB IOVE I NINE ENN I MEFTON

435 PST--FNSGAMPGANNRI DSQ LNVKHKRIKISFQ LDEK FVY P E W I G S E E N K N I R K L A I G S L 4

DB 444 IRIYGYPIGLLSAIS-FDAIRNGLALAAPALGQDVGYHF-----VPRLAGGV 4

493 PSNERYWLDPGTPOVTLKEDSVNVEFSLKLYLNSVNSLSFVGDSLYLEFGISE---LPSLW 2

Db 492 SSP-----RGANGNIFLGSAL-TWGTNGNGNELDITKW 3

550 YSFPRLSDLTALNQVKTDIEASST---DNGTTTNGTTTADTSSGS-----TGAG 5

Db 522 H--SPAVIED-----APTTFITVNSSGVLQNSGSQQTSTPMPNSNGNESIPYRWNTNSY 5

QY 600 TGNNTNTSQTIVSNPTLNTYRSFCIDSKPTSAKIDETNWADPNVI--EAR1YAEY-----6

Db 574 DYN5VRFALISKPAGNTKQ--VESLETTALKLDLTN-SLPNKFTQENNIFESYAMLDC 6

QY 653 ---RLGIONE---IPITNAGNFIKNTIGVGFTSTGSRVLRASNG-----DQRPTG 6

Db 631 RQWSLGT RKDSTWLTNT INNFYNTQQQLASTAGENANPRNILNALTAKGFRDIDG 6

QY 700 NFQPFY-----VEGYLGYYQTRTGFW--YGTYYKLNNSPYDVLDSPRVGTEYNQFR- 7

Db 691 NVD-ILYSNNTNKFTYY-YQVGGAITTWPEVQVNYKTSANITYNLTTRDFGSTPATQD 7

751 R T S L T Y P V M G G Y L T E E G A R S - - F S N T P Y I R A Q G D T P E S R S I F O S G Y S D N T Y E Y I O S V L G F 8

Db 749 ANTVSSKLNGAYLSTGDQGWYNGSIYKASFTSSQ-----GYWODEKGLTTTASN 803
 QY 809 DGRNNLVGVKASSPLNSNPNGLNEMIAATYLRSGIARFTSGLPNOOPFGTHOV 868
 Db 804 AVISMWTKAGY-----SIRPDD-----YFVSYSKI-----Pf----- 831
 QY 869 ISVSPGDQFSSIKNIRTFPGNOJMYELFTNNENKSSVYTLRLADSNPDASSSPSLSL 928
 Db 832 -----EKETIAVANYRSDS-----YYVOLNGETSVNTVAVY---SPSSA----- 869
 QY 929 IDVNEIGVLLPLDLSFYVNAAGNALFSSNPGSPSYAVTFNFNINSLDIAFEGSGAK 988
 Db 870 LTLNPKRIINPLMNRD-----NVIGQAFISRN-DIPSSF-----FENKINDIV----- 912
 QY 989 YTSDFMGCTIOFQDEYLIONGFTSOVARNFVYTNOSFLNSLVDTTPA--NAGTNRVYVDP 1046
 Db 913 -----TTEADGTEVLDISKYINSIRYRTPPONNPHIRLRLLVID 950
 QY 1047 DGNLTNOLNPLKYQIYLDGKYYDAKLNNNLVT--FSTNNFGALPSSWVYPAIGSTGLIL 1105
 Db 951 RSRATNDFIKLLPQV-LVDEYVAVPOANSVFSVSDQETGEDALPGYLVPAISIPILII 1009
 QY 1106 AIMTILGAIPIPLRAORIKODGKFTFKYDVTLLAAVGSYKRIITQTA--NKKKPA 1163
 Db 1010 ALALALGIGIPIPSORNMKIKQFALSNKKVDILTLAVGSVEKIOINTSTVNIKRPQ 1069
 QY 1164 ALGAGKSGDKKPAAPAKPAPKSPAPKSPAKPSPKSPGAPTKPLAP--KPAAPKPT 1220
 Db 1070 MLOA-----NKKDGA-----SSPSKSPAPAKPPTGPKP--SAPGAKPTAPAKKAPAPT 1118

RESULT 4
 A36968
 p1-like adhesin precursor - Mycoplasma p1rum
 C:Species: Mycoplasma p1rum
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-Dec-1999
 C:Accession: A36968
 R:Tham, T.N.; Ferris, S.; Bahraoui, E.; Canarelli, S.; Montanier, L.; Blanchard, A.
 J. Bacteriol. 176, 781-788, 1994
 A:Title: Molecular characterization of the p1-like adhesin gene from Mycoplasma p1rum.
 A:Reference number: A36968; MID:94131957; PMID:8300531
 A:Accession: A36968
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1144 <THA>
 A:Cross-references: GB:L19685; NID:9404770; PIDN:AAC36866.1; PID:9404771
 A:Genetic code: SGC3

Query Match 5.3%: Score 337; DB 2; Length 1144;
 Best local similarity 21.5%: Pred. No. 5.5e-10;
 Matches 299; Conservative 168; Mismatches 491; Indels 434; Gaps 74;

QY 5 KKLK---SYTLIGLAVFAGLGSASFGKOSDKSNNTOLVNO-----ARTLDANSYRL 55
 Db 2 KIKKNYKYLILSLISTIVSAALISYSTFNKQDISPIINONVKSFSNPISVGNKVGK 61
 QY 56 AGLGONGSLFNTLVLDVDFNFTANGTIIKIDSTFKPLGLD-----LSPD-----CGCY 106
 Db 62 IRHMNNNN--FNGVEIKNGGFVYLSTQATRIDAFGNILWFEPEKIASDSQYANLAK 120
 QY 107 KVKQIVSDYTSNRPFDOQOTAYYALIV-----NDEANVHLKRIINTNSRIGNRNNNSK 161
 Db 121 KVEITQDEGDSN-----IKLLILPKNTPKQASIDPKLVAYNELTGSKSQK 172
 QY 162 FVI-----GGVNDPAHVR--FTDGTFR--NFTQOTGELVNFILDAFLPK 206
 Db 173 ATVVQIENNVNLXGSGTWPSPTIKLHIDPKKVVADYPMQWSSQSS-----STFTK 227
 QY 207 DLHPDWNLXIQRIKLPN-DVNT-----AVYPM-----VGRVSGTNADGMFDC 250
 Db 228 EDHPSWYVANNSEKVEHNDADQNTQYNGLKSANMVLPMKQIYITNIGNFAKNGVLIIFG- 286

QY 251 NGGOITNTDPIAQT-----KTTDNONPSTFNSGAMPGA-----NNRYSQLVN 294
 Db 287 GNGSLYN--DPEALSIGMKKLDLKPYSQNDNNO-----ANGGIPAYLLRLYDPS-- 338
 QY 295 KKRKTSFQDLDRINTNSRIGNRNNNSKFVIGVDNPAHYIFETDDGTKEFNTQTOGE 354
 Db 339 KPLDGS-----APNR--RWNQSYAFIGQDNEFTYVPRLANGVOINASTD----- 382
 QY 355 IYNDFILDAPLPKDLHPWMLYI-----QRKILPNVNTAVPMVPGRS 401
 Db 383 -----BATYLYLAAGITVQAKESQAREVISNS--NT-----S 413
 QY 402 GTNADGMFDCGGOITNTDPIAQTCTTNDONPSTFNSGAMPANNRYSOLNVKRIK 461
 Db 414 TNKVVYTIQKRSLDOLTGAWTINTKOT-----AANSIDAL-----LFG 453
 QY 462 TSFQDEREYVPEMTGSEENKNTRLATGSLPSNERVYLDIPGTPOVTLKEDSVNFSR 521
 Db 454 TAFNIDSLINLP--TKLENLFTIQ--NVQYESYF--DVGATMSVSSAVGYIYFBDK 506
 QY 522 -----LYNSVNSLSFIDGSIYIFGTSELPSLMYSFPTRLSDTLALNQVKTIDEASST 576
 Db 507 KNHASSSTDINTYTASNGWNMLGRYAF--WSYK-PN--NDIGSIPOKPTND----- 555
 QY 577 DNGTTNGTTTADTSSGSGTAGTNTNTSQTVSNPTLNTFY--RSFGIDSKPTSA--NK 632
 Db 556 NNNAI-----YGNLSLILLEN-AIYVYSTLSFGSLKUGLGLTK 594
 QY 633 IDETNADPNVYIARVAYRGLIONEIPITNACNFRITIGVGFSTGSRVLRASYN 692
 Db 595 IEMPSKEN-----ENTICYAM-OVQKSIYVLINEBK 625
 QY 693 GDQRPNGNQPFLVYRGVLYGQOTRGTFWYG-----TYKLINSPIYLDSPVGRTEJNO 748
 Db 626 SDRSLAYHGPSISIG-----ESNLVGSAKYQDMQPYVKINNSNIGYVSD--YSNTTN 680
 QY 749 FRATSLYPMWGYLTEEGARSESNF--PYIRAGDTPESRSIFQSGYSDN-----TYETI 802
 Db 681 IINGVAI-----YVT--GIKQENDTIPITASQFE-----IGNSPEEDNSTTKYTGTL 727
 QY 803 QSVL-----GPDGIRNLNVGVKASS--FLNSNRPNGLKEMIAATYLRSGIARTSGLPN 858
 Db 728 OPTIPWADFQGL-NSTNFNSEISLMLNNQRTKNNE----- 764
 QY 859 QOPFGTHQIVSVSPGQFSSIKNIRTFPGNOIY-----YELFTENN-----KSSYXT 908
 Db 765 -----HFIVTKSP--EISB-----YCGNMIWTEREYVYNGSSNADWKSGKRAWE 808
 QY 909 IRLADSNPDASSFSPTSILDVNEIGVILPLLDNSFPTV-----N 949
 Db 809 VK-----DSNLSNSSTTVGWQVIGLSDNLADSYTVQKNNEDQPGDFVILRTRDDT 860
 QY 950 AAGNVALPS--SNPSPG--SYTAVTFNOMLSLIAFGSGAKYTSDFMGCTIOFKPDEYLI 1006
 Db 861 OKNNDIFFQJINNTREPGISYCKL--KQ-----YGSFYETLSIDPLSL 905
 QY 1007 QG--FTSOVARNEFTVN-----OSFLNSLVDEF--PA 1034
 Db 906 GNGQFVNNLSNOILTNLNLMLLVQVNLSTVGNPLDSKSTIRYKNGFLNEVQVKNPT 965
 QY 1035 NAGTNR-----YVDDGMLTNQNLPLKVOIYLDGKITYDAKLKN-----NLVTF--YNN 1085
 Db 966 IEGSTPYGVYIVASNVDFVSQSAFTYASWNSLTQYDWPRTNSSKINVINNSLIFAG 1025
 QY 1086 FGALPSWVVPFAIGSTLIGLAIMITIGLAIPLAORIKODGKFTFKYDVTLLIANG 1145
 Db 1026 FSAADWILPLVIAIPVIVALLITIGCSIGIPMAKHKAIKVGPELOHDGVGLTSAVG 1085
 QY 1146 SYTKIIT--TOTAVKKKPAALGAGKSGDKKPAAPAKP--AAAKKSPAKKSPAKP 1202
 Db 1086 GVFKKIDTNTNSNVKSKPQML-----KAAAKKPTVTPARSQTLNDSVSRPPTS 1136
 QY 1203 SGAPTKPTAPKP 1214

Db 1137 S-----APKP 1141

RESULT 5

LUTMAP

adhesin PI precursor - Mycoplasma pneumoniae

N:Alternate names: cytohesin PI; hypothetical protein 116a

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 31-Mar-1989 #sequence-revision 30-Sep-1991 #text-change 07-Dec-1999

C:Accession: S03725; A38791; J00288; S04516; S62807; S73339; A60102; A27597; P10008

R:Su, C.J.; Tryon, V.V.; Baseman, J.B.

Infect. Immun. 55, 3023-3029, 1987

A:Title: Cloning and sequence analysis of cytohesin PI gene from Mycoplasma pneumoniae.

A:Reference number: S03725; MUID:88057593; PMID:3119495

A:Accession: S03725

A:Molecule type: DNA

A:Residues: 1-1627 <SU1>

A:Cross-references: EMBL:M18639; NID:g150166; PIDN:AAA25424.1; PID:g150167

A:Experimental source: strain M129-B16

A:Accession: A38791

A:Molecule type: protein

A:Residues: 60-77 <SU2>

A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated

R:Inamine, J.M.; Denny, T.P.; Loebel, S.; Schaper, U.; Huang, C.H.; Bott, K.F.; Hu, P.C.

Gene 64, 217-229, 1988

A:Title: Nucleotide sequence of the P 1 attachment-protein gene of Mycoplasma pneumoniae

A:Reference number: J00288; MUID:88297153; PMID:2841195

A:Accession: J00288

A:Molecule type: DNA

A:Residues: 1-1627 <INA>

A:Cross-references: GB:M21519; NID:g150138; PIDN:AAA8325.1; PID:g150139

A:Experimental source: strain M-129, ATCC 29342

R:Kenzel, R.; Hermann, R.

Nucleic Acids Res. 16, 8337-8350, 1988

A:Title: Repetitive DNA sequences in Mycoplasma pneumoniae.

A:Reference number: S04516; MUID:88335593; PMID:3138660

A:Accession: S04516

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1187-1413 <MEN>

A:Cross-references: EMBL:X13087

R:Hilbert, H.; Himmelreich, R.; Plagens, H.; Hermann, R.

Nucleic Acids Res. 24, 628-639, 1996

A:Title: Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S62797; MUID:96177562; PMID:8604303

A:Accession: S62807

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1156-1199, 'GAT', 1203-1253, 'VRRS', <HIL>

A:Cross-references: EMBL:U34795; NID:g1215683; PIDN:AAQ43678.1; PID:g1215685

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Hermann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73329

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1627 <HIM>

A:Cross-references: EMBL:AE000002; GB:U00089; NID:g1673651; PIDN:AA95661.1; PID:g167365

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

R:Jacobs, E.; Gerstenecker, B.; Mader, B.; Huang, C.H.; Hu, P.C.; Halter, R.; Bredt, W.

Infect. Immun. 57, 685-688, 1989

A:Title: Binding sites of attachment-inhibiting monoclonal antibodies and antibodies for

A:Reference number: A60102; MUID:89138634; PMID:2465270

A:Accession: A60102

A:Molecule type: protein

A:Residues: 237-246/702-708 <JAC>

C:Comment: The protein is the major adhesin mediating the attachment of the mycoplasma to

A:Genetic code: SGC3
C:Superfamily: adhesin PI
C:Keywords: cell adhesion; membrane protein
F:1-59/Domain: signal sequence #status predicted <Sig>
F:60-1627/Product: adhesin PI #status experimental <Mat>

Query Match 5.1%; Score 328; DB 1; Length 1627;

Best local Similarity 21.3%; Pred. No. 2, 8e-09;

Matches 268; Conservative 147; Mismatches 404; Indels 442; Gaps 68;

QY	251	GNQGIT	---NTDPIA-----QYTTTDN-----QNPSTFNSGAMPGANN	286
DB	506	GNQHYTKSAHTAPLSIGFVRVRYNATGATVGMPEALLFGSMVKKQTDGLKDLDPFN	565	
QY	287	RYSQSLNVKHKRTSPQDERINTNSRIGNRNNKFFYIGG---VDNPAHYRFPD--	341	
DB	566	RW-----FEYVRAVAVAGAKFVGR---ELVLAGITMGDRVPRLLYDEL	608	
QY	342	GTRFNFNTOTGEIVNDFIL-----DAPI-----LPRDLH--PDMYNL	377	
DB	609	ESNLNLVAGQGGLLRDDQLFTPYGMANRPDLPIGAMSSSSSSSHNAPYFHHNPPMDQR	668	
QY	378	YIORKITLNDVNTAVVMPVGRVSGTNADD-----GMFDGNGQITN--TDPIDQT	426	
DB	669	PID-----NVAFAFIKPEW---DKNGKDAKITYPRYSGMW---AMQVYMSNKLTDQ	716	
QY	427	KTTTDD--NONPSTFNS-----GAMP-----GANN-----RYSQSLNVK	457	
DB	717	PLSADVENENAVOPNSLFPAILNPELLALPDQKVKYKGENEPANERYENOKLIVAPIQ	776	
QY	458	-----HRIKTSFO-----LDEKEFYVPEMTSEENKNTIRLATGSLPSNERVWIL	501	
DB	777	GTRWMSHFSPILSFSGFNLVGSVDQVLDYVWINGRYG-----NNHR-----	822	
QY	502	DLPETQVTLKEDSVNFRILYNSLSFT-----GDSITIFGT	542	
DB	823	---GVDDITAPQTSACSSSGSISTNTGSRSLFTPEFNIGVGKLANVOATLGSGQTMITGG	879	
QY	543	SELPSLWYSSFPRRLDLALNOVKT-----DDIEASTDNGTGTGTTATSGSGTS	597	
DB	880	S-----PRTLDQANL--OLMTGAGCRNKKASSGSGSDENHTFTSATGMDQGGSG	928	
QY	598	AGTGNTTNTSO---TVSNPLNTYRSFGIDSKPTFSANKIDETPMAD--PNTVIRIAYEY	652	
DB	929	TSAGNDSLSKQDMSISGDSLTQDGNALDQO-----EATYTNLPRLTPT---ADW	978	
QY	653	RIGIONEIPITNAGN-----FIRNTIGVGFTSRSRYVLRASYNGDQRPFGNPFEL	706	
DB	979	---PALSFNTKNNNAQRAQLFLRGLGSI-----PVLNRSGSDSNKFOATDQKWSY	1027	
QY	707	VFGYLGQQRRTGTFWGYTK--LLN-----NSPYDVLDSPRVG---TETNG	748	
DB	1028	T--DLHSDQTKLNPATVGEVGNLLNPALVETTYGNTRAGSGSNTTSSPQIGKIPDQNN	1085	
QY	749	FRRTSTYTP-----VMGCTLEEGARFSFNTPIYRAQGD-----	782	
DB	1086	DSKATLITLPELANTPDQVGLVYSGTTVSFQGLGMLV-----FTTDPYKPPAAGVLGLD	1138	
QY	783	-----TPESRSIFOSGYSDNTVEYIOSVLFQDGT-----RNNLVGVKAS	822	
DB	1139	LTFGLDASDATORALIMPRPMAAFRCGSMV--NRIGRVEYSVDLKGVMADQAOSDSQSTTT	1197	
QY	823	SFLNSNRPNRNGL-----EMIAATTYL--RSQIGLARNSGL-----	856	
DB	1198	ATRNALPEHPNALAFQVSVESAVYKPNITSSGGTQSTNSSPYLHLVKPKVQTSQDKLDD	1257	
QY	857	-----PNO-----OPFGTHOIVISVPDQSSIKINIKITFPGNOLMYFLTNEENKS	904	
DB	1258	LKNLIDPNQVTRKLQSGFDH---STQPPQ--SLKTTTPVVG-----TSSGNLS	1303	
QY	905	SVYTLRLADSSNDPASSFSPTSLIDVNEI---GVILPLDLSFTYVNAAGVAFSSN	960	
DB	1304	SVLS-----GGAGGGSSGSGQGVLDLSPVEKVSGLVQLPST-----SDGNTS--STN	1351	

QY 961 PGSPGTYANVTENQ-----NLSDIAFEGGSAKITSDEWGTIOFKPDEYLLONGFTSQA 1015
 1352 NLAAPT-----NIGNDVYGVGRLS-----ESMAAKMNDVDGIVRTPLAELLDEGQATDG 1403
 Db 1016 RNFVT-----NOSTFLNSLV-ETPAN-----AGINRYV-----VYDP 1046
 QY 1404 POSVKKSPDQIDFRLTHFVLDLPVMTLYDYDYLPLFDIPASVPPKAVRLKYSF 1463
 Db 1047 DGLTQNONPLKVOIYLDGKYDAKLNN-----NLVFS-----YNNFGA 1088
 QY 1464 DTN--PQSLGLRLEFPKP--QDTQPNNVQVPPNNGDPLPLTASOGPQLTSPFQ 1518
 Db 1089 LPSKVPYATGSLGLTAMITLGLAIGPLRAOKIDGKFTTEKVDTLFAVSVY 1148
 QY 1519 WPDVPLPALTIVIVIVLSVTLGLAIGIPMKNOALKAGFALSNQKAVLTKAVSVF 1578
 Db 1149 KRIITOTANVKKKPAALGAGSGDKKPAAPAKPAPAPKAPKASPAKPTGKSGAPRK 1208
 QY 1579 KEIIMRT-GISQAKRL-----KQISAAPKAPAPPPVPPKPPQAPKPPVOP---PKK 1625
 Db 1209 P 1209
 QY 1626 P 1626

RESULT 6

A:14180
 A:adhesin p1, group 2 variant precursor - Mycoplasma pneumoniae (strain TW 7-5)

C:Species: Mycoplasma pneumoniae

A:Variety: strain TW 7-5

C:Date: 10-Apr-1992 #sequence-revision 04-Sep-1998 #text-change 07-Dec-1999

C:Accession: A14180; S73339

R:Su, C.J.; Chavoya, A.; Dallo, S.F.; Baseman, J.B.

RefSeq: Immun. 58, 2669-2674, 1990

A:Title: Sequence divergence of the cytoadhesin gene of Mycoplasma pneumoniae.

A:Reference number: A14180; MUID:9031696; PMID:1973413

A:Accession: A14180

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 194-376; 377,855-931; 944-1160 <SUA1>

A:Note: neither the complete nucleic acid sequence nor the complete translation are shown

A:Note: strain TW 7-5 may be ATCC 29085

A:Accession: B41480

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 563-665; 676-678; 1208-1210; 1256-1258; 1385-1387 <SUA2>

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.

RefSeq: Immun. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73339

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-217; 'GLPQRTESGNTST', 234-243; 'NAADTRAKSNKLGAEAG', 265-269; 'GSQGS', 274-278; '360-362; 'A', 364-368; 'KDLPMKS', 376-383; 'D', 565-676; 'N', 678-854; 'F', 856-914; 'EN', 916-919; 'KAT', 1099-1119; 'TV', 1122-1135; 'V', 1137-1208; 'N', 1210-1256; 'T', 1258-1633 <MIN>

A:Cross-references: EMBL:AE000002; GB:U000089; MIM:91673651; PIDN:AD9561.1; PID:9167365

A:Experimental source: ATCC 29442

A:Note: this sequence is used to show identical portions of A14180 not shown in that ref

C:Comment: The adhesin p1 from group 2 strains differs markedly from the adhesin p1 from

9342, see PIR:S73339.

C:Genetics: The protein is the major adhesin mediating the attachment of the mycoplasma

A:Gene: PI

A:Genetic code: SGC3

C:Superfamily: adhesin p1

C:Keywords: cell adhesion; membrane protein

F:1-59/Domain: signal sequence #status predicted <SIG>

F:60-1635/Product: adhesin p1 #status predicted <MAT>

Query Match 4.8%; Score 310.5; DB 2; Length 1635;

Best Local Similarity 20.4%; Pred. No. 2,4e-08; Indels 595; Gaps 82;
 Matches 324; Conservative 168; Mismatches 503; Indels 595; Gaps 82;
 QY 41 LVNQARTLDANSYRLAGLQNGSLPNTVLRVDNFTTANGTILKIDSPTRIVGLDS 100
 Db 219 LPNQGSSSSGSDST---NOTGAMFGLKVKD-----ATVDSKOSTELKCESS 263
 QY 101 DDCGKYVQIYSDYTSRRRPROR-----ATVDSKOSTELKCESS 263
 Db 264 -----SSSTSTSTTQKGGSSNENKVALQAVAKKSGSGSGDGTQY 311
 QY 133 LLYNDEANVHLKRIINTSNRIGNRNNKFKVIGVD---NPAAVIRFTDGTKEFTN-Q 188
 Db 312 LESNDLANAPTKR-----GSNNQOVLKADDFGAPESSSGTQDGTPTPMPWL 208
 QY 189 TQGEIYND-----FILDP-----ILPKL----- 422
 Db 363 TTEQIINDPAKFASTILLYDAPYARNRTAIDRYDHLDPKAMANTPPSMRTPKNNHGL 228
 QY 209 -----HPDWYN-----LYIQRI---LPNDVNT----- 482
 Db 423 WDKKARDVILQITGFENPRHPEWFGQYVADNEKTFGVDSSEMTKQGFQKQKADSDKS 267
 QY 229 AVVPPW-----VGRVSGTNADGMDGNGQIT---NTDPIA-----QTKT 541
 Db 483 APIALPEAVFANIGLITWFGALLVFG-GNGHVKSAHTAPLSIGFVRVYANATGTSAT 327
 QY 268 TDONPSTENSGAPGANNRYDSQLNWKRIKITSFOLDERINTSNRIGNRNNKFKVIG 594
 Db 542 VTGWPFYALFEGMV---NKOTDLKMLPNNRNKWEFYVRAVAVAKAFVGR---ELVLA 365
 QY 328 G---VDNPAHVRFTTD---GTFENFTNNOGELVNDFFIL-----DAPL----- 654
 Db 595 GITMGDAIVAYRLLYDLESNLVLAOGGLRDLQLEFTPYGNANRPDLPLGAMSSSS 408
 QY 366 -----LPKRLH---PDWMLYIQRLKLPNDVNTAVVPPGVRSGTADP-----G 408
 Db 655 SSSHNAPIYFHNPPWQBRPID---SYDAFTKPE---DKNGDAKIYFYSRG 705
 QY 409 MFDGNGQITN-IDPIAQTKTTD---NONPSTFNS-----GAMP-----GANN 448
 Db 706 MW---AMOVYNSNKLITDPLADVVENAAQPSLPAIINPELLAALPDKKYKENE 762
 QY 449 -----RYDSQLNVK-----HRIKISFO-----LDEKFIYPPMTGSEENK 822
 Db 763 FAANEYERFNOKLTVAPTOGTMSHPSPLSFSGFMVGSYDQVLDVPMIGNRY 822
 QY 483 NITRLATGSLPSNERWIIDPGTPOVTLKEDSVAVFSLYNSLSLPI----- 533
 Db 823 G-----NNHR-----GVDDITAPQTSAGSSGISTNTSGSRSLPTFSNIGVG 865
 QY 534 -----GDSIYIIGTSELPISLWYSPFRLSDLTALNOYKT-----DIEASSDNG 579
 Db 866 LKANVQATLIGSQMTIGGS-----PRTILDQANL-OLMTGACRNDKASSGQSD 914
 QY 580 TTTNGTTTADYSSGSGTAGCN-----TTNNTQTSVNSPRLNTRYSGI 623
 Db 915 DHKFTLSAGMGQEQESGTSAGNPDLSLKDKIKSGSLTQOGNMEQOEATNTNLP 974
 QY 624 DSKPT-----SANKIDETWN--- 638
 Db 975 NLTTTADWPAALFTKNNKNOAROLFRLGLGSLPYLNGSGQDNKFFAEQOKMSYTD 1034
 QY 639 ---ADPNVTEARIYAEIRLQIONEIP-----TNAGNFIRNTIG---GYCF 681
 Db 1035 IQSDQTKLINLPAYEVN-GLNPALEYIEFGMTRASGSGSNTSPGIGFIPQOSTNT 1093
 QY 682 GSRVYL-RASVNGDORPTGN-----FO---PFIYV-----GYLQY----- 714
 Db 1094 TSKAVLITPGLAMTPQDGNLVVSGTSFSLQGLMLVTFDTLKPRAGLIQLTGLDAS 1153
 QY 715 -OTRGFTMYGIYKLLNNSPYDVDSRV-----GTEINQFRTSLITPVMGVLTBECA 768

Db 1154 DATORALIM-----APRPMWAFRSGMVMNLRGHVESVMDLKGVMADQ--A 1195
 Oy 769 RSESNRPYIRAGDGT-PESSRIFQSGYSDMTVEYIOSVLGFDRIKNNLVNG--VKASS 823
 Db 1196 QSSQSSQSTTATRDALPE-----HPNALAFQYSVSEASVYKPPETSSGQTQSTNSSP 1246
 Oy 824 FLNSNBNP-----PNGLEMIATATYLRSGTGLARGLPNOFGTTHQVIVSPGD 875
 Db 1247 YLHLVAPKPKVIGSDKLDLDDKLNLDPNQVTKL-----RQSGTDH---SQQPP 1293
 Oy 876 QRSSTINRTIPFGNOLMTFLPENNKKSSVYTLRLADSNPDASSFSPTSLIDVNEI- 934
 Db 1294 Q--SLKTTTPVFC-----TSSGNLSSVLS-----GGGAGGSGSGSGSGQSDVLSPYE 1337
 Oy 935 ---GVILPLLDNSFTYVNAAGNVALFSSNPGSGSTYAVTENG-----NLSDIAFEGSG 986
 Db 1338 KVSQMLVGQLPST-----SGGNTS--STNNLAPNT-----NTGDVYGVGLS-----ESNA 1382
 Oy 987 AKYTSDFWGTIOQRPDEYLIQNGFTSQVARNFVT-----NOSFLNSLVD-FPPAN- 1035
 Db 1383 AKMNDVDGIVRTPPLAEELLDGEGQTADTGPQSVKFKSPQIDFNRLFTHVPTDLFDPVTM 1442
 Oy 1036 -----AGTNRV-----VYDDPGNTNQLPLKVOIQIYLDGKYYDAKLKN- 1076
 Db 1443 LVYDQYIPLFIDIPASVNPMMVRLKYLSEDTN--EOSLGLRLEFFKPD--ODTOPNNV 1497
 Oy 1077 -----NLVTFSS-----YNNFGALPMMVVPALGSTGLAIMIILGATGIPL 1119
 Db 1498 QVNPNNGDPLPLTASSQSGQTLFSPNOMPDYVLPALIVPVIYVILSTGLAGIIPM 1557
 Oy 1120 RAQRKLQDGFETTKFVKVDTLTAAGSVYKIIITQTAANYKKRPAALAGSGDKKPAAAA 1179
 Db 1558 HKKQALKAGFALSNGKVDVLTAKAGSVKEIINRT-GISQAPKRL-----KQTSAA 1608
 Oy 1180 KPAAPAKPSKPAKSSPAKPPGKSGAPTKP 1209
 Db 1609 KPGAPRPVPPKPGAPKPPVQP----PKKP 1634
 RESULT 7
 A30588
 N:140K adhesin precursor - Mycoplasma genitalium
 N:Alternate names: attachment protein Mgpa
 C:Species: Mycoplasma genitalium
 C:Date: 08-Jun-1989 #sequence, revision 01-Dec-1995 #text_change 20-Apr-2001
 C:Accession: A30588; J00090; B64221; A61605; S18721
 R:Ballio, S.F.; Chavoya, A.; Su, C.J.; Baseman, J.B.
 IntAct: Immun. 57, 1059-1065, 1989
 A:Title: DNA and protein sequence homologues between the adhesins of Mycoplasma genitalium
 A:Reference number: A30588; M0ID:89173298; PMID:2925238
 A:Accession: A30588
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1444 <DIAL>
 R:Imamine, J.M.; Loachel, S.; Collier, A.M.; Barille, M.F.; Hu, P.C.
 Gene 82, 259-267, 1989
 A:Title: Nucleotide sequence of the Mgpa (mnp) operon of Mycoplasma genitalium and comparison with the sequence of the adhesin gene
 A:Reference number: J00090; M0ID:90060815; PMID:2583522
 A:Accession: J00090
 A:Molecule type: DNA
 A:Residues: 1-1444 <INA>
 A:CROSS-references: GB:M31431; NID:g150157; PIDN:AAA25420.1; PID:g150159
 A:Experimental source: strain G-37 (ATCC 33530)
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhmann, J.; Nguyen, D.; Usterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.; C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; M0ID:96026346; PMID:7569993
 A:Accession: B64221
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1444 <TIGR>

A:CROSS-references: GB:U39696; GB:I43967; NID:g1045869; PID:g1045876; TIGR:M6191
 A:Experimental source: strain G-37
 R:Mader, B.; Hu, P.C.; Huang, C.H.; Schillz, E.; Jacobs, E.
 Zentralbl. Bakteriol. 274, 507-513, 1991
 A:Title: The mature Mgpa-adhesin of mycoplasma genitalium.
 A:Reference number: A61605; M0ID:91321682; PMID:1863319
 A:Accession: A61605
 A:Molecule type: protein
 A:Residues: 59-68 <MAD>
 R:Peterson, S.N.; Schramm, N.; Hu, P.; Bott, K.F.; Hutchison III, C.A.
 Nucleic Acids Res. 19, 6027-6031, 1991
 A:Title: A random sequencing approach for placing markers on the physical map of Myc.
 A:Reference number: S18693; M0ID:92051396; PMID:1945886
 A:Accession: S18721
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 106-176 <PER>
 A:CROSS-references: EMBL:X61522
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1991
 C:Genetics:
 A:Gene: Mgpa
 A:Genetic code: SGC3
 C:Function:
 A:Description: this protein plays an important role in the attachment of the respect
 C:Superfamily: adhesin p1
 C:Keywords: membrane protein
 F:59-1444/Product: 140K adhesin #status experimental <MAT>
 Query Match 4.48; Score 285; DB 1; Length 1444;
 Best Local Similarity 20.98; Pred. No. 4.4e-07;
 Matches 270; Conservative 159; Mismatches 477; Indels 386; Gaps 68;
 Oy 24 SASRFGKQSD-----KSNMQVLVNAQRTIDANSVRLAGGONGSLFNTLRDQDN 75
 Db 418 NTSPEFVGVDTHDKKQGFKNSSPIALPPEAYFANIGNVAIG-----NSV----- 464
 Oy 76 FITAANGTILKIDSTFKPLGLDSDCGGVKQYVSDYTSRNRFPDQRTAY-VAL- 133
 Db 465 FIFGNGHAKMT-TNPL-----SIGVFRK-----YT-----DNFSKSSVGMPPAVL 508
 Oy 134 ---LVNDEANVHLKRTFNSNR-----IGNRNNKSFVIG--VDNPAH 172
 Db 509 FGLINPQTN-GKLDPLGTNRWFYVPRMAVSGVKKVGN-----QLVLAGTLMGDTAT 562
 Oy 173 VIFRTDQGR--ENFTNQGEIYNDIILAPILKDLHWDWNLVYQRIILPDVNTAV 230
 Db 563 VPRUKYDQLEKHNLVAQGGGLREDLQITPY-----GMAN---RPDIP-----V 605
 Oy 231 VPPVGVSGTNNDDGFMDCGNGQITWTDPIDIAQTKTTDQNPSTFNSGAMPAN--NR 287
 Db 606 GAW-----LQDER-----GSKFGPHYFLN 625
 Oy 288 YDSQNLVYKRIKTSFOLDERINTNSRIGNRNNKSFVIGVDNPAHVIFRTDQ---T 343
 Db 626 PDIODVNNN-----DIVEALISSYKMTDKL-----KHVYPRYSGLYAMQ 665
 Oy 344 KMFNTNO-TOGEIYNDIILAPILKDLHWDWNLVYQRIILPDVNTAVPVPVGVSG 402
 Db 666 LFNWNSNLTNTPISANVNSYAPNSLF-----AAILNEDLTGL----- 706
 Oy 403 TNADDGFMDCGNGQITWTDPIDIAQTKTTDQNPSTFNSGAMPANRRYDSQNLVYKRIK 462
 Db 707 ---SDKFFYKKEHFEANEDRFNQLSLNPNP-----NTMAAYLVNVOGFTT 752
 Oy 463 SFOLD-----EKYV-YPEWTSSEENKNTRLATGSLPSNERWILIDIGTPQVTLKEDSVN 517
 Db 753 GPNLDSSTPQDFLDFLWIGN--GKPSNSPSPSTASAS-----STPLPFPFSNTNVG 802
 Oy 518 VFGRLL--YLVSVNSLSIGSIYIFGSELSPLW---YSPFTRLSDLALANQVKDD- 570
 Db 803 VKSMITQHLKKEINR-----WFLIPNPSDITWAGYVQSANOKNGIPFEQVPSNN 855
 Oy 571 ---IEASSTNDGTTNGTTTADT-----SSGSTAGTGTNTNTSQTAVSNPTLN--TYR 619

```

Db      856 STPEPNSDDKVTSPSSGSSKFTYYPALPNSISPTSDMINALFTFNK--NNPORNOLLR 913
      620 SPGIDSKPTSAKIDEIN-----WADPNVIEARI--VYAEYRLGIONEPIITNAGN 667
      914 SL-LGTLIPVLINKSGDSNDQFNKDEQKMDKETETNEGKLPQFEVFN--GLINALLHTYGF 971
      668 FLRNITG---GVGF-----TSTGSRYVLKASVNGDORPTGN-----FQPLPYFGY 710
      972 FGTNINSIDPKIGFADSSSSSSSTLVSGSLNMTSODVGNLVYINDSPFGOLGQMTTF 1031
      711 LGYQRTGTETWYG--TYKLINSPDYDLDSRVRGFTNQFRSTLTFYPMAGYLTLEGA- 768
      1032 TDEIRPRTG--YLGITLSLQDQTLIMADQ-----WTSFK-----GSYLDSDGTF 1075
      769 -----RSFSNTPYIRAGDTPESRSIFQSGYSDNTYETIYQSVLGEFGIRNNLVYKAS 822
      1076 KSLMDPTALKSLPNSSTYTDINPLTSPFOLYQPKVKAYQTTNY-----KLLPEVDAT 1131
      823 SFINSNRPNNGLEMTAATYLRSQLGLARTSGLENOQPFQTHOIVSVSPGDQFSSIKN 882
      1132 SAAT---NMTSLIKLITTKNIKAKLGKSTASSQGNNGGVQSOTINTITTTGNISBGLK 1187
      883 IRTIFPGNOLMYFLFTENNKSSVYTLRLADSSNPDASSFSPSLIDVNEIGVILPLLD 942
      1188 EETSIQAEFLAKKFFDSKQNNKS-----EIGI---GD 1215
      943 NSFYVNAAGNVALFSSNPGSPSYTAVTNFNONLSDIAFEGSGAKYTSDFMGTIOFRPD 1002
      1216 STFKMPD--GKLTGVVSTP-----LVNLIH-----GOGASDSD--TEKISFKRG 1256
      1003 EYLQNG--FTSQVARNVTNQSFLNS-----LVDFTPA--NAGTNRYVYVDDGNILJNQN 1054
      1257 NQIDENRFLPIPLVELEDPNMFYDQVPLVNLSEGFQDASIRKVTI-----SYSVEN 1311
      1055 LPLKVOIOYLDGK-----YDADLKNNNLVTFESYNGFALPSMVPYTAIGSTLGLTAIM 1108
      1312 QTLGVRLFKDPQFOQPIPLVNLASTGPOVFOFQFNMA---DYVPLAYTVIVVILLS 1368
      1109 IILGTAIGIPLRAQKLDQKFTTFRKVDTLTAAGSVYKKIITQV--ANVKKKRAALG 1166
      1369 VTLGLTIGIPMHRNKKALQAGFDLSNKKVDVLTAKVSVKEILNRTGISNAPRK----- 1423
      1167 AGKSGDKKPAAPAKPAKSPASSPAKP 1198
      1424 -----LKAATPKPL--PK--TPPKP 1440

RESULT 8
T30852
Outer membrane protein A - Rickettsia conorii (fragment)
C:Species: Rickettsia conorii
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30852
R:Gillmore Jr., R.D.
Gene 125, 97-102, 1993
A:Title: Comparison of the ompA gene repeat regions of Rickettsiae reveals species-specific
A:Reference number: Z20904, MID:93194085, PMID:7680636
A:Accession: T30852
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1029 <GIL>
A:Cross-references: EMBL:L01462; NID:g152485; PID:g152486; PIDN:AAA9908.1
C:Genetics:
A:Gene: ompA

Query Match      3.98; Score 247.5; DB 2; Length 1029;
Best Local Similarity 20.99; Pred. No. 2,0e-05;
Matches 269; Conservative 131; Mismatches 468; Indels 411; Gaps 62;

      7 LKSTLLIGLAVF--GALGSAFGFKQSGDKSNDTQLVNQARTLDANSVRLAGLQNGSLF 65
      7 LRTVVGAGIATLEGAIIKAT-----TTKLLNMASTVLTINV----- 43

```


Dn DAFNLKNNVNNNSNTVFYLNCA-----TTWKTAGTGVFTQDYGINSVLVENQTTPFLAG 955

Oy 859 QDPFGTHOIVIS--SPGQFSSIKNIRTIPGNOL----- 892

Dn 956 ANP--TSNSVYGEFKTSGAEMGLVIGCVKANQIDITGTIRSGNGAKGGATLVNA 1013

Oy 893 -----WFFLETFENKNKSSV-----YTRRL----- 911

Dn 1014 QERLIANLANLNDKRAGLONSMMNFIVNNGNLNTNANFSQOTPHGFENLKANNITWDKG.1073

Oy 912 -----ADSSNPDASSFPSPTSLID-----VNEIGVILPLIDNSF--ITYNAA 951

Dn 1074 SVSGCGNFGVDNMAANANAVIKRVNFSDDGTLIYGGESAGNSLTLENMFNSYNINAK 1133

Oy 952 GNVALFESSNPGSPGSY-----TAVNT-----FN--ONLS 978

Dn 1134 AONLIENNNSNFGSAYSFNDTKNVTEKGTNTLINSDPSRKLGSVIDNNSIFNERDLT 1193

Oy 979 D----IAEFGCARKY-----TSDFWTGIQPKDE-----YLQNGFTISOV 1014

Dn 1194 DKTYTLLISGNITKNINMOALADNVFSKNLMDLIHYDGEGTLIRTDNNITYFO--ETQSN 1251

Oy 1015 AHENFYTNOSFINSLVDFTPANAGTRYKV--VDPPG-NLTNONPLEKVOIQYLDGK 1067

Dn 1252 GQKFVEEP-----FNGSITRYKYPTTIHSSPHTEADSKDIWNQ--VRKQPFPIGK 1301

Oy 1068 -----YDKAKTKNNNLV--TFSYN-NEGALPSWVPVPAIGSTGLIATMIIGLAIG 1116

Dn 1302 TPVCYGVCCVIARYKKQDLIGSSAPFAMSLNFA-----TVVGST-----LLIG----- 1342

Oy 1117 IPLRAQRKLDDKGFRTTKKYVDTL 1140

Dn 1343 ---SAQEKANNNGGISIMFGKNML 1363

RESULT 11

S76412
hypothetical protein slr0408 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

R:Accession: S76412
R:Nameko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda Res. 3, 109-116, 1996

A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*.
S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4199 <KAN>
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BA18541.1; RID:g1653604
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: *Synechocystis* hypothetical protein slr0408

Query Match 3.6%; Score 233.5; DB 2; Length 4199;

Best Local Similarity 20.6%; Pred.No. 0.0011;

MATCHES 284; Conservative 161; Mismatches 525; Indels 409; Gaps 66;

Oy 16 LAVFAGASASFC-----FKSQSKSDNTQL-----VNGARTLLANSVR 54

Dn 531 VSILGLIGRIAGFPFWTOTFAASSQNNGGLEANNVTASVDGFLILINLAQFLERN--- 587

Oy 55 LAGIGONGSLFTVLRDVDDNETANGTITIKLDSFRKPLXGLDLSDDCGYKVKQIVS- 113

Dn 588 ----LSLNPDTFTV-----VTNANQI-----STLPFGVVQVQDOSSNSSEVILRL 631

Oy 114 ----DYTSNRFDQOTRAYALLVNB--ANVHLKRIINTSNRNIGNNNKSFEYIGV 167

Dn 632 EOELPYTT-----VNGCFENIIVK--YSGNQLGN-FTSSPVALAOSP 670

Oy 168 DNPAHVIRFTDDGTFKNFT--NOTQGEIVNDITLARILPKDLHPDWMYLYIQRIILPNDV 226

11-11-21-0408 - *Synechocystis* sp. (strain PCC 6803)
RESULT 11
S76412

hypothetical protein S11040
C;Species: Synechocystis sp

```
A; Variety: PCC 6803
C; Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
```

C/Accession: S76412
P/Name: Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, M.; Tanaka, A.; Watanabe, A.; Yamada, M.; Yasuda, H.

N. Kato, S. Shimpo, S. Takeuchi, C. Wada, I. Watanabe, 1996
O. K. Okumura, S. Shimpo, S. Takeuchi, C. Wada, I. Watanabe, 1996
109-136

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Cyanothece*

S.
A; Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S76412
A;Status: preliminary

A: Molecule type: DNA
A: Residues: 1-4199 <KAN>
PID: g1653604: PIDN: BAA18541.1; PID: g1653604: PIDN: BAA18541.1

A; Cross-references: EMBL:D90915; GB:AB001339; NID:g105007; JGI:JGI-105007. The nucleotide sequence was submitted to the EMBL Data Library, June 1996

C; Superfamily: Synechocystis hypothetical protein S10408

Query Match	Score	DB	Length
3.6%	233.5	2	4199
30.6%	0.0011		

Best Local Similarity	20.00;	Indels	409;	Gaps	00;
Matches 284; Conservative 161; Mismatches	525;				

16 LAVFAGLSASF-----FKQSDKSDNTOL-----VNOARTLDANSVR 54

531 VSLGLIGARIAGFTWTOTFAASSSQNNQTLGLEANNTASVDGFLTLINLAQFLEPN--- 587

55 TACIGONGSIFNTVLRDVDNFI TANGTIIKLDSTKPLYGLDSDDCGGYKQIVS- 113

00 UNOXXXXXXXXXXXXX :||| | ||::: |
 01 DEFEVT-----VTNANGOI-----STIPVEGVVQTDSQSNNSEVILRL 631
 02

Db 588 ---LSLNFIDEIYI
---LSTOPOBPAVYALIYNDE--ANVHLKRINTNSNRIGNRNNSKFIGV 167

OY		---	DYTTSKRREDOQKAI INLEVNOZ	:	: ::	:	670
			UNCOBESNTIVR---YSGNOLGN-FTSSPVALOSP	:			
114	---	DYTTSKRREDOQKAI INLEVNOZ	:	: ::	:		

```

Db      632 EQEIPYTT-----VNGQFFGNLAVN      226

```

QY 168 DNPAAHVIRETTDDGKTENFT - NOTOGEIVNDFILDAF ILFADENFTOMINZ.....

```
Db      671 DLIIYINPT-SGTNDYONNOQIITAFNGCLDPIIPE-----NOLIMFAVTNNS 723
QY      227 NTAAYPWPGRVSGTNADGMDCGNGQITNTDPIAQTITTDNONP-STFSGAMPAN 285
Db      724 NOAILITDVA-----VNOSRIYLTILASTAIPEIYVQYNPDAN 764
QY      286 NRYDSQILNVKRIKTSFOLDERINTNSRIGNRNNNSKRVIGVDNPAHVRFTDGTGF 345
Db      765 S--DQOLKRAQ-KITLAFVTSNGTAPITAMGLINTFAGVINQTPV-LSMLTSD----- 816
QY      346 NPTNQT-----QGEIYNDFIIDA-PILPKDLHPWNLXIQRKILPNDVNTAVVPP 396
Db      817 -FAQDTSFALATISQGDILLAMSSDTPITPISVLAEGDYLL--VFADNLKNDSANPPS 873
QY      397 VGRVSGTNADGMDFCGNGQITNTDPIAQTITTDNONP-STFSGAMPANNTYDQ 453
Db      874 NSQFTIKTSD-----GNTTPTPNVSLAONTITLTLTNSVNASQIVESYLSGTMLTJN 927
QY      454 LNVKRIKTSFOLDERFVPEWTGSEENKNTRLATGSL-----PSNERWILD 502
Db      928 LVLATATNTSFWV-----PDFTNSVQASGSTAPTSLGVSISLITLITPNTLNVQ 981
QY      503 IP--GTPVYTLKEDSVNVSRLYLSVNSLSFICDSI-----YIGTSELPSLWYSFPT 555
Db      982 IFNGGQFYTVANGNSTSPI-----TVTSVTVADTSITLVLNQLIGQGLVITTY--TPN 1033
QY      556 RLSDDLALNQVKTDOLEAS-SINDNGTITNGTITTA-----DISSSTGA 598
Db      1034 SGENNNLVSNCTPQTVASFSTNNELTTASGTGYKTAFAFSPGDSGSSITTIPTQTTI 1093
QY      599 GTGNTTNTSQTIVSNPTL-NTYRSFGIDSKPTANKIDETNMADPN-----YI 644
Db      1094 NSDVAVATLYQNSTALQNVAVMVDTSALSKITPGQNYG-PNPAALITTAQOOSDY 1152
QY      645 EARIATVRLGIGONEPITNACNFIRNTIG-GVGFSTGSRVYLAASVINGDORPGENFP 703
Db      1153 YVLEPGDNQWGLAARFFSSQPOGDKVTLGVPNGMLLAWMLTOLDSQDPMTTQIAT 1212
QY      704 FLYVEGYLGQOTRTGFYGYTKLLNNSPYDLDSFVGTETNOFRSTL----- 754
Db      1213 F-----NOTTM-----TMP-TILGANGINPNSELSISSINGOPAI 1250
QY      755 -----TPYMGYLFEEGAR 769
Db      1251 FWTESRPPSYNLVSEONPLVYLRGLSELGTVINQGLSVANGTYSTAGTYLGGVAL 1310
QY      770 SFSNT-----PYIRAG-----DIPESRIFQSGSDMTYEQSVLGFDIRN----- 813
Db      1311 EMTNTNTGDFNPVAVLEFNGGGITINSFVSV--QGSVEFWFKLPSTDGVGLANLAGVF 1368
QY      814 NLNVGKASSFLNSNPNP-----NGLEMA 839
Db      1369 DSLINDESLITLNNNSNPQISGYITTGMMHYVGVYDVKOILDXLDQALVNTLENTA 1428
QY      840 ATTYLRS-QIGLARTSGLPNOQPFQTHQVIS-VSPGDQFSIKIRITIEPQNL-----W 893
Db      1429 FANLPQSGTLTLGLSSGYSYLDDEFAYNSILSYDNGSSPSSNNFNLMTGSLNGAIW 1488
QY      894 YLLFTNENKSSVYTLRLAD--SSNPDASSSF-----SPSLIDVNE-IGVYLLPL 941
Db      1489 -----GVNEVGSYHQAFFEPYTAGPETVYVWSDSSGNSWOSPVSNINPDEVVPTLLSAA 1543
QY      942 DNSFYTV--NAAGNALFSSNPGSGSYTAVN-----TFNQLSDIAFE 983
Db      1544 NNPIMDIYASNPAGANNNAQIAPNGNDPIIEQVNLIGQOGESEITGFTVTTSSNQLMVGCD 1603
QY      984 GSGAKYTSDFW--GTIQKRPDEXYLQNGFTSOVARNFVTNOSFLNSL-----VDTTPANA 1036
Db      1604 GTGNMVF-SESQOLGYI-----LAENADSTIPOLEFISGDKLILNLCANGATFSHRVMA 1655
QY      1037 GNNRYVVDPCGN-LTNQNLPLKVOIYLDGKYDDAKLNNNLVTTSYNNFGALPSPVVP 1095
```

```
Db      1656 TETFTIEVDEGSPLTS---PATVNI-YLQGO-----TDPITFT--SLSPINOGCP 1701
QY      1096 TAIGS-----TLGLAIITLGLAIGPLAORKLDQKRTFKYVDLTAAVGSV 1147
Db      1702 VASNSPDYLDNOVLTGIATKEAND-----ASLSVDSGF-----VIDTNPALIAAV 1747

RESULT 12
A:97859
190-Kha cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #text-change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fourminter, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MID:21442074; PMID:11557893
A:Accession: A97859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AL03811.1; PID:g15620410; GSPDB:GN00173
C:Genetics:
A:Gene: rompa

Query Match
Best Local Similarity 3.6%; Score 232.5; DB 2; Length 2021;
Matches 278; Conservative 147; Mismatches 536; Indels 365; Gaps 64;

QY      9 SYTLIGLAVFALGSASFQKSDKSNQNTQVNOARTLANSVRLAGLQNGSLNTV 68
Db      162 NYTGILNIALGA-----NAALIIQSA--PAKITLAG-NINGGIIITV 202
QY      69 LRVDVDFITANGTIKIDFTKPLGLDLDGCGYKVKQIVSDYTSRRRFQQR 128
Db      203 KTD-----AALNGTGTNTALAVYVVGAGIA-TLEGALIKATTTKLTNAASVLTITNVN 255
QY      129 AYYALLVNDNAVH-HLKRINTN-----SNRIGNRN-----NNSKFVIGVDNPAHYI 174
Db      256 AVLTGAIDWTGTGVNDGVNLNGLALSQVGNIGNTNALATISVAGAKATLGAVIKATTT 315
QY      175 RFTDGTGRNFNTNQ--TQGEIYNDFIIDAPILPKDLHPWNLXIQRKILPNDVNTAVV 232
Db      316 KLTDNASAVTFNPVVYVTAIDN-----TGANNNGIYT 348
QY      233 WP-----VGRVSGTNADGMDFCGNGQITNTDPIAQTITTDNONPSTF-NSGAMPGA 284
Db      349 FTGDSITVYTNIGNITNA-LATISVAGAKATLGAAIIKATTTKLTDNASAVTFNPVVYVTA 407
QY      285 NNRYSQILNVKRIKTSFOLDERINTNSRIGNRN-----NNSKFVIGVDNPAHYIR 337
Db      408 ---IDNTGNANNNGIYV-FTGDSITVGN---IGNTNALATISVAGAKATLGAAIIKATTTK 460
QY      338 FTDDGKREFTNQ--TQGEIYNDFIIDAPILPKDLHPWNLXIQRKILPNDVNTAVV 395
Db      461 LTDNASAVTFNPVVYVTAIDN-----TGANNNGIYV 493
QY      396 P-----VGRVSGTNADGMDFCGNGQITNTDPIAQTITTDNONPSTF-NSGAMPAN 447
Db      494 TGDSTYTGINGINTNA-LATISVAGAKATLGAAIIKATTTKLTDNASAVTFNPVVYVTA 551
QY      508 QVTLKEDSVNVSRLYLSVNSLSFICDSIYIFGTSLSPLWYSFPTLSDIPLANOVK 567
Db      590 KATLGAIITKATTTKLTDNASAVTFNPVVYVTAIDN-----NTGN 629
QY      568 TDDIEASPDNGTNTGTTTADTSGSTGAG-----TGNTNTSQTIVSNPTL 615
Db      630 ANNGIYTFNGNSTVIGNIGNITNATLAVNVGAGIATLEGAVIKATTTKLTNAASVLTITNV 689
```

QY 616 NTYRSPGIDSKPTSAKIDETNMADPNVTEARITAEYELGIONIEPIITMNGNIEFRNTIG 675
 690 NALVTALD-----NTTGVDNVGVNLNGALSGVLTGNIGNTNLALITSVAG-KATLGG 742
 QY 676 VGFSTGSAVVLRAAS-----YNGDQRETFGNF--QPLFY 706
 Db 743 AVTKATTTKLTLDNASAVFTTNPVVVTGAIDNTGNANNGIAFTFGDSVLTGNIGNTNLAT 802
 QY 707 VEGYLGYOQTRGTGTEWYGTGKLLNSPYVDLDSPPV-----GTE 745
 Db 803 VNVGAGGLLRVOGVKASNTINLTLDNASAVFTTNPVVVTGAIDNTGNANNGIYVTFGDSV 862
 QY 746 TNGFRRTS--LYTPVAGYLTEEGARSFNTPYIRAQGDPE--SRSIQSGISDNITYYI 802
 Db 863 TGNIGNTNLALITISVCGKATLGGALIKATTTKLTLDNASAVFTTNPVVVTGAIDNT 918
 QY 803 QSVYGFEGTIRKNNLVGVKASSFLNSRPNPNGLEMIATYTLRSOIGLARTSGL-PNQP 861
 Db 919 -----GNANNGIYVTFGDSVLTGNIGNTNL--ALVNVGAGVLTQAGGSLDANNID 967
 QY 862 FGTTHQVIVSPGDQFESSIKNIRTIFFGNOLMYFL--FTNENK-----SSVYTLR 910
 Db 968 FGARSTLEFNGPLDG-----GGMNAPIYFFGALANGNALLNVNFKLTATHT 1016
 QY 911 LADSS--NPQASSEF-----SPTSLIDVNEIGVILPLDINSFYVNAAG--NVALESSNP 961
 Db 1017 IGTVAEINIGAGNFALDASAGDVTILMQDHFRLDLSALVSLNLTGVGVNNILAAVL 1076
 QY 962 GSPSYTAVNTFENQL-----SDIAFEGSAGKATYSDQWCTIDQKDEFLIONGFTSQVA 1015
 Db 1077 VAPGVDEGTVEFDGCVGLNIGSNVA--GAARIGDVGN--KNTILLITNAYVITTD 1130
 QY 1016 RNFVYTNQSEFL-NSLVDFTPA--NAGT-----NYRVVDPDGNLTNLNPLKVQIQYLD 1065
 Db 1131 VNLBEIQVVLINNNADFTSTAFNAGTIOINDAYTIT--DANNGLNINPAG-NKF-- 1183
 QY 1066 GKYYPAKL-----KNNNLVTFESYNNFGALPSSVVPPIAISTGILAIMIIT--LGLA 1114
 Db 1184 -AHAAQALILONSSGNDTITLGN-----IDPP--NDDEGIYILNSVTAGKKLTITIA 1232
 QY 1115 IGIPRAQRKILODKGK-----TFEKKVDTLLTAAGS-----YKKIIT-- 1152
 Db 1233 GKRTFGCAHKLODIFVEKGEDGTAGTTEFTNTITVLDTQGLEIATTAANVLFKQAVOL 1292
 QY 1153 TOTANY 1158
 Db 1293 TOTGNT 1298

RESULT 13
 F82885
 hypothetical protein U0482 [imported] - ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: F82885
 R:Glass, J.I.; Isfakowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 Submitted to Genbank, February 2000
 A:Description: The complete sequence of ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: F82885
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-4688 <GLA>
 A:Cross-references: GB:EO02145; GB:AF222894; NID:g6899476; PIDN:RAJ30894.1; GSPDB:GN001
 C:Experimental source: serovar 3, biovar 1
 C:Genetics:
 A:Gene: U0482
 A:Genetic code: SGC3

QY	2	NISKKUL---	SYTLIGGLAVFAGALSASFQOSKSDNDNTQVNOQAATDANSVRLAGL	58
Db	2027	HISKNEFKLEQNIINGEVAVSYLITLNS---	SLIVNNKNVILN--FTLD-----	2070
QY	59	GQMSLEFNVLRVDNDFITLAAAGTITIKIDSEFTKPLDGLDSDDCGKRVQVSDYTT	118	
Db	2071	---NKSNOILYRLVDYYIIDDNDNTINDKNKPKPANNVTRIIDLAPG---	FTTIS---KS 2121	
QY	119	RNFEDQROTBRAYALLVNDDEANVHLKRINTSGNRIGNRNNNSKPFYIGVDNPAHVIREPTD	178	
Db	2122	NNMTNTSTSSQSEFEVINSD-----	DGNEY-----LDNLFIATISPKK 2158	
QY	179	DGT-----KRFNTNOGELVYNDFILADLPEDLHPDWNYLTIORKIL--	PNDVNTAV 230	
Db	2159	GQTLLEPVKNVITIKQN-----	NKYLIGQI--TNLEPE--NRVYESTILAKPKTKRPL 2209	
QY	231	VPMPFVGVSGTNADDGDFDCGNGOITNTDPIAOTKTTTNDONPSPFNSG-----	AMPAN 285	
Db	2210	V-----VELINKDISFOTOAGNY-----	KVIOIRK--SONPSTVDRKORIKIKLDGIG 2255	
QY	286	NRV--DSQNLV-----	KHRIKTSFOLD-----ERINTNSR 314	
Db	2256	NAMMEKOLEITYSANDNSTYTAITIKLEKNLEFELTINENKRRYTFPKIELINDNMTK	2315	
QY	315	IGNRRNNS--KFVY-----	GGVDNPAHV-----IRFDDGTRKFNFTQOGE 354	
Db	2316	TPFNKSDSIDOKPFIYLSNNGVGNITIEIODRVDNHLNSAKIRFELDNDLVLSDEQAT	2375	
QY	355	IV--NDPFLDAPIL-----	PKDIAHDWNTIKYORKILFENDVNTFVMPFVGVSGTNADP 407	
Db	2376	IYNNNOFTTSKAVITDNOKYLEATFSLVNLKDTIINKLEFNTRKRNASKNGINDTVN	2435	
QY	408	-----	-----GMPCGNG 415	
Db	2436	IYDATNLILINDKLITGPLEHLEKEFEANKNTINISVLELDTNNHISKNYLFIAKPSNDG	2495	
QY	416	QITNTDPIAOTKTTTNDONPS-----	TENGAPG-----ANNRYD 451	
Db	2496	QSVLNPPLAADKIVTNNNNKKELFNLNLITNSKROITFPGLYVNSQNTINDENKKE	2555	
QY	452	SQLVNKHRIK-----	TSFQLEKFEVYEMT-----GSEENKNTI 485	
Db	2556	KSNVDVKITVAPRTTITIOKNGNMTFPPQPNAKQKFEFNISDENVDFTDIDATITSSDH	2615	
QY	486	---RLATGSLPSNERYW-----	ILDIIPGTPVYTLKEDSV-----NVFSLYLSVNSLSFT 533	
Db	2616	DTKKITTTKLOKQDNQWYLEDTISDAINDYTKLDDISITPRMAFANLKIQLEN-----	2670	
QY	534	GDSIYIEFGRSELPSLWYSEFTRLSDLLALNOVKTDIEASSTONGTTNGTTTADTSS	593	
Db	2671	-----KEQISFTTQSGPQLVITSISHDLMMNDQPNAS-----	MOOTTAAYS- 2712	
QY	594	GSTGAGTGNNTTSGTVSNPTLNTYRSFC-----	ISDKPTSNANKIDETMW--ADPNVIEAR 647	
Db	2713	-----GVDNLYNNRKI--	KLYEKRNSNOQKSLVESNELITQK--DQYQIFETPLISVANR 2763	
QY	648	IYAEVRLGID-----	NEIPIITNAGN-----FIRNTJGGVFSTGSRVYLRA 689	
Db	2764	QYSKEKIKITSNNNFELTNSGTVNSPNNVNASKTOIYIDNTINATINITSATII	2819	
QY	690	SYNGDQRPNTGNFQPLVYGYLGYOQTRGTWGYGYKLLN--	NSPYDVLDSPPRQVTEHTNQ 748	
Db	2820	-----	TYLKSADHFPQYQDSITITLYLKSND 2844	
QY	749	FRRTSLTPYMGGLTEGAR---	SESNPTIYRAGQDPRBSRSIFQSGYSDNTYTYI--- 802	
Db	2845	DERQBIESTYKTITSISSDGEATVFSFTDQTLKEBAN-----	YLLIVG 2886	
QY	803	---OSVLAGFDGIRNNLNVGYKASSFLNSRPNPNGLEMTAATTYLRSOIGLARTGLDLPNQ	859	
Db	2889	FVRKPTLATYATINNNDAN-----	NVIEDNNSYN-----FTLIVDHKIVNVSNN 2933	
QY	860	QPFQTHQVIVSVS--PGQOFSI--	KNITIPFGNQ-----LM--YFLFTNNENK 903	

Db 2934 DSINTTQTVNIDIDIGIQRTWINKKIOLVYTSNDGEELISDQKTLIMANNHSEFLSNLK 2993
OY 904 -SSVYTLRLADSSNPASSF-----SPTSLIDVNEIGVILPLIDNSF----- 945
Db 2994 HNRKTKTLKVRRIINDNKTSLIFHLKNGIADMEYVKTKOSISISISIEPIARAKNLOST 3053
OY 946 ---YTVNAGNVALESPNSPGSYTAVTNFONLS---DIAFEGSGAKYTSDFMGTIOF 999
Db 3054 QTRFILDNDPNVL---SN-----EEEAIIINGENLSKAKKIVYS-QYLEFTFNFLAL 3104
OY 1000 KDEYILQNGFTSOVARNVYTNOSFLNSLVDTFPANAGTNYRVVDPDC---NLINQNL 1056
Db 3105 MOD-----TINSITPINK-----PIKAATN--IGDINSKRIYNNFNINP 3143
OY 1057 LKVOIQLDGKYDAKLKNNNLVTFPSNFGALPSPVNPVPAIGSTLGILAIMILGLAIG 1116
Db 3144 LKIDNNF---SVIGPQOSDNRKRTILEYN-----KTNVKNV 3177
OY 1117 IFLRAQR-LQDKGFTFKKVDLTAAVGSYKKIITQATANYKK 1161
Db 3178 LELKANDILQNLXYFATFKQDNNTV-----KSIIVSYSIKKE 3217

RESULT 14

AB2018
hypothetical protein all1696 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2018
R:kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MID:21595285; PMID:11759840
A:Accession: AB2018
A:status: preliminary
A:molecule type: DNA
A:Residues: 1-1599 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAF78062.1; PID:g17135516; GSPDB:GN00179
A:Experimental source: strain PCC 7120
A:gene: all1696

Query Match 3.5%; Score 223.5; DB 2; Length 1999;
Best local similarity 19.8%; Pred. No. 0.0013;
Matches 306; Conservative 162; Mismatches 502; Indels 577; Gaps 74;

OY 27 FGFQSDKSDNDTQ---LVNQARTLDANSVRL---AGLQNGSLFNTVLRVDN- 75
Db 278 FVFTQSGNDNTDSDADPLTGQAQIIEINPNEFNRTIDAGIYRKASLDGEVFKDANNNG 337
OY 76 -----FITANGTIKLDSTFKLYGLDSDCGGKRYKQIV-SDY----- 115
Db 338 IODAGEVGVQYVELINPTNGDIATTT-----TNSGGYQFGSLTPGNGYQVR 387
OY 116 -----TTSNREFDQOTRA-----YVALLVNDEAN-----YHLKRI----- 146
Db 388 TAPGTGIIFSTANGSDNDELSDANPSTGLTQITLISGEPNGILDGVLPLASLGNFVE 447
OY 147 --NTNSRIGNRNNNSKFVIGVDNPAHY--RFTDDGKFNFTNTOGEIVDFILDAP 202
Db 448 DKNANGIQDAGETGIGNATVNLDAENIATATDONGLYSFTNLRPGYKQYFV----- 503
OY 203 ILPKDLHPDMYNIYIORKLIPND-----VNTAVVWPVPGRVSGTAAODGMF----- 248
Db 504 -----QPNGFNQVSPQNGGDAIDSDGLVSDIVMLSPGE-NDTIVVDAGFYTASLDGF 556
OY 249 -----DCNGSOI-----TNTDPIAOTKTTTNDQNPSTFNSGAMP----- 282
Db 557 VFKDANNNGIQDAGEVGVQYVELINPTNGDIA-----TTTNSGCGYQFSGLTGNGYV 613

OY 283 -----GANNRYSQNLNVKRIKITSFOLD----- 305
Db 614 RFTAPGTGIIFSTANGSDNDELSDANPSTGLTQITLISGEPNGILDGVLPLASLGNFV 673
OY 306 -ERINTNSRIGNRNNNSKFVIGVDNPAHY--RFTDDGKFNFTNTOGEIVDFILD 362
Db 674 FEDKANNGIQDAGETGIGDGTAVKLLDTSNGVIAATTTTDDNGLYSFTNLRPGYKQYF 733
OY 363 A---PILPD-----LHPDMYNI-----YIORKILP-----NDVNTAV 392
Db 734 SGFDGISPKNNGGNDALISDGLISDVYNIISPENDDTVAGFQTSASLDGPFVNDVN--- 790
OY 393 VPMVPGRVSGTNADDMPCGN---GOIT-----NTDPIAOTKTTTNDQNP----- 434
Db 791 -----NNGIQDAGEIGVGAYVELINPANGDIA-----TTTDDANGYGFSG 834
OY 435 -----ESTFN--SGAMGANNRYSQNLNVKRIKITSFOLD----- 468
Db 835 TPGEVKRTAPAGVNFSLANGNNDALSDANVSTGITQITLISGEPNGILDGVLPL 894
OY 469 ---KFVPEMTGSEENKNTIRLATGSLPSNERYWLIDIPGPQVTLKEDSVNFSRLYL 524
Db 895 ASLGNFV-----EDKN-----ANGIQDAGEIG-----IGDQTVLLDTSNGV----- 932
OY 525 NSVNSLSFTQDSIYIEFTSELPRLWYSF--PTRISDLTALNQVKTIDIEA----- 573
Db 933 --IATTTTDDNGLYSF--TNLRPGDYKVGQVQPNGFNGVSPQNVGNDALISDGLISDV 989
OY 574 -STDNGTTNG-----TTTAD-----TSSGTCG--GT 600
Db 990 LSGENDTIVDAGFYKTASLDGFVNDVNNNGIQDAGEVGAVSVTLTGCGGGLIINGI 1049
OY 601 GNTTNSQTVSNPTLN-----TYR-----SFGIDSKRPSANKIDETRMADPN-- 642
Db 1050 GDTVTYTTTNNAGNVPFAELTPGQYQVGVSGLPAPFOQTQAVGANDVDSD-ANPSTG 1108
OY 643 -----VIEARIYAE-----YRLGIONE--IPTN 664
Db 1109 KTQVTVLVSGENNLTLTDAGIYONAGDLSTKTDGLTNTVPGQITTYTIVARNNGLLTAVN 1168
OY 665 A--GNIRMTIGVGTSGTSGRVYLAASVNGDQRPNGNQPLXYVGILGYOQTRGTGM 722
Db 1169 ALVSDIIPSNLTNTVTSVASSGATDNOASG--TGNINDHVTLLTGSSITTYVTGTV 1224
OY 723 YGTYKLLNNSPYDV-LDSPRVGTE--TNOFRRTSLTPYMGVGLTEEGARSESNT----- 774
Db 1225 SNAASGSSTRFPDQNSPLDGTGNTTRFTQGTIVT-----ARAFSRVDGTNG 1274
OY 775 -----PYIRAQDTPESRSIFQSGYSDNTVEYIQSYLGEDG----- 810
Db 1275 AMSAAVLYSGTGLGVTDS-----EGNGGNNTHHVVDNNGGRDNYVLQFSEAVVLDAK 1329
OY 811 ---IRNNLVGVKASSF-----LNSNRPRNGLMEIATTT-----YLRQI 848
Db 1330 LOYVLKDSDISVWIGNFNNPLTNLSDSTLNSFCGEYANVTSSSDRMADVNAAGVNGNT 1389
OY 849 GLARTSGLPNOQFFGTHQIVSVSPGDOF-----SSIKIRITFPNQQLMYEL 896
Db 1390 VIAALD-----TDTSPEDNFKRLRLDLVKALIPALVNTATITTAPEG--FT 1432
OY 897 FTNENKSSVYTLRLADSS-----NPDASSFSPTSLIDVNEIG----- 935
Db 1433 DTPNPNNSATDWTIVAPSDGLSITKTDGLTNTVPGQITTYTIVARNNGILITATNALVSD 1492
OY 936 VILPLDINSFTYTNAG-----NALRSSN-----PGSGSYTAVNFTNQLSDIAE 983
Db 1493 IIPSNLTNTVTSVASSGATDNOASGTGNINDHVTLLTGSSITTYVTGTV---VSNAAST 1549
OY 984 GSGAKYTSFEMWTIQKFPDE---YLIONGFT-----SOV-ARNFVTNOSFLNS----- 1027
Db 1550 GSSTRF--DFDGSPLDGDIDGNRTFTQNGITYTAAPSRVSDGTNAMSAAVLYGSIYTGGL 1607
OY 1028 -LVDFTPANAGTNYRVVVDPDGNTLNQNLPLKVOIQYLDGKYYDAKLK-----N 1075

Db 1608 GYTSTSENGNGNNH-VVDNNGGRDNVYLQFSEAVVLDKAVLYLVKDSIDSVWIGNFN 1666
 QY 1076 N---NLVTFSYNNFG-----ALPSWVYPTA---IGSTLGLAI 1107
 Db 1667 NPLFTLSDSTLNSFGFYEAANYTSSSDRMADVNGVNTIYAL 1713

RESULT 15

A64556
 toxin-like outer membrane protein HP0289 - Helicobacter pylori (strain 26695)

C.Species: Helicobacter pylori

C.Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C.Accession: A64556

R.Tomb, J.F.; White, O.; Kellavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.

A.Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A.Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: A64556

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2893 <TOM>

A:Cross-references: GB:AE000547; GB:AE000511; NID:92313377; PIDN:AA007355.1; PID:9231338

Query Match

Best Local Similarity 3.5%; Score 223; DB 2; Length 2893;

Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;

QY 14 GGLAVFG-----ALGSASFGRKOSKSDNTQVLYNOARTIDANSVRLAGLNGSL 64
 Db 287 GGYTFNKEFSATNTNTAFSSGFSNFKGVSSFN-GTFSNASTYFD-MQATTONSSFNCGT 344
 QY 65 F-----NVLRDVDNFTTAANGTIIKIDSFETKPLYGILSDDCGKYKQIVSD- 114
 Db 345 FTTNNOTNPTNNAHQIIONSSFSGNATLTK-----GFWNFQAFNNS 387
 QY 115 ---YTTSRNRFDOGRTRAYVALLVNDQANVHLKRTINSNRGNNSKFPYIGV---- 167
 Db 388 NHOITTONASFNNATFNNTGKITIEKDAFEN---NTFNTSVDTNNMS--VTGGVTTISG 441
 QY 168 -----DNPAHYIRFTDGTGFNFN-----QTQGEIVNDFILDABI--LP 205
 Db 442 KNDLKNGSTIDPSSSKITLA-QCTTFNLTLGSEKSVTILNSSGGITTYNLNHAINGLT 500
 QY 206 KDLHPWYLYIQRIKILPNDVNTAVVPMVGRVSGTNADGMEFGNGGITTDTPIAQTK 265
 Db 501 SALKTN-----ESLSNPOSFAQGLMDIITYNGV-----TGQLN-ENNAISK 541
 QY 266 TTTDNONPSTFNSGAMPGANRRYSOLANKHRI-KTSFOLDERINTNSRI-----GNRN 319
 Db 542 PTSSSPKSSSTNSTQV-----YQGYKIGDTIYKIQETFFSHNSIILQALBSGYT 591
 QY 320 ---NSKFEVIG9-----VDNPAHVIR--FTDDGTFENFTNOGEIYND--- 358
 Db 592 PPPIYNGSKFDLSASNTINADMPYDHKYYIPKSONFTESGTYLPSVOIMGYSYTSFRQ 651
 QY 359 -FIIDAPFLPKDLHPM--YNLYIQRIKILPNDVNTAVV-----PMVGRVSGTNADGME 410
 Db 652 TFSANGSNLYIGVNTDHNWSSSGTSGVSGSALNGHCGPMWYQCTGTT--NGTY 709
 QY 411 DC-----NGCQITN-----TDPJAQTKTTTDNONPS-----TFNSGAMP 444
 Db 710 SAHYVITANLRSGNRIGTGAANLIFNGVDSINIANATITQHNAGIYSSSMFTSOSMD 769
 QY 445 GANN---RYDSOLNVKHKIKTSFQLDKRYEYPEWTSSEKNKTRLATGSLPSNERYWI 500
 Db 770 NSQNLNGLNSNGKLSYGTFTTNEAKDKGKLFENAGQAVFENTFN----- 814
 QY 501 LDIGTQVYTLKEPSVNVFSLYLN-----SVNSLSFGISDIYTFGTSSELPSTLW 549

Db 815 ---GGSYQ--FSGDSLNFNSNNQFNSGSEISAKNASFNNANFNNSASFNEFNNSNATTSF 869
 QY 550 YSPPTRLSDLTALNQVKTDDIEASTDNGTNTTGTATDTSAGSAGAGTNT----- 603
 Db 870 VGEFTNANSNL-----QI-----AGNAVEFSTNSQNTANFNNGVNSISGATFDDNV 919
 QY 604 ---TNTS-----QTVSNPT--LNTYRSFG-----IDSKP---TSA 630
 Db 920 ENGPTNTSVKGOVTLNNTITKLNLNPLSFGDGTTFNNAHVINIAESITNGNPITLVSSS 979
 QY 631 NKIDETNMADPNVIEARIYAERYRLQIONEPITNAGNFRITRTIGVGFTSGSVYLRAS 690
 Db 980 KELEYNNAFSKMIQOLINOQH--GASSEKLVSSAGNGVYDV-----YS 1022
 QY 691 YNGDORPTGNQPEL-----YFEGYLG-----YQGTTRG--TFWGYTK- 727
 Db 1023 FNNQ---TYNFOEVSQNSISIRRLGVNVPFYVDMKSDHLYQNALGFMTYMPNSYNN 1079
 QY 728 ---LLNNSPY-----DVLDSPRVGTETNOFRRTSLTYPVMGGYLTEGARSFSNTPYIR 778
 Db 1080 NLGNMANTLIYYDKSIDFYASGKTLFTKAFFSQT-----FTGQNSAIVFGAKSIWT 1130
 QY 779 AGCDTPESRSITQSGYSNT-----YE-----YIO----- 803
 Db 1131 SLSDAQSQSTIIIFRG--DNKGAGSNDASGHCNNLOCIGTTHGYEAKIYITIGSTESGMR 1188
 QY 804 -----SVLGFEDGIRN-----NLNVGVKASS--FLNSNRPNNGLEMTATTYLRS 846
 Db 1189 ISSGGGASLNFNQLGILLTNATLYNRAAGTQSSSNFTS-----NSANIOAQNSYFID 1242
 QY 847 QIGLARTSLPNOQPGTTHQVLSVSPGQFSSIKNIRTIIPGQNLWYLFTENNKSS- 905
 Db 1243 D--TAONGGNPN--FSEFNLNIDFNSNFRGVGKTQSVKFNKNAISFTNSNLSG 1297
 QY 906 VYTLR-----LADSNPDASSFSPSLIDVNEIGVI-----LPILDN 943
 Db 1298 LYQWQAKSVLPDMSNLVSVG---TSSIKAMALNSQNASINASHSTLEOGDNLVNDT 1354
 QY 944 SFY-----TVNAGNAL--FSSNPGSPGTYT---AVN-----TFNQLSDIAFEGS- 985
 Db 1355 SSLNLNSTINVSNNATINDYASLIASNGHLNFGAVNFNSANITTSLNSSIFYKGAIV 1414
 QY 986 --GAKY-----TSDFGCTIOFKPDEYLIQNGFTSQVARNFYTNQSFNLSDVFTPANNG 1037
 Db 1415 SLGQFNLSSNSLDPGS-----SATSMTAFNFYDN-AFSOSPITFHQA--- 1459
 QY 1038 TNYRVVVDPPGNTLTONLPLKVOIQYIDGKYIYDAKLKNNNLVTFSSYNNNGAL 1089
 Db 1460 LDIRAPLSLGGNLNLPN-----NSSYLD-----LKSOLV--RGDQSL 1496

Search completed: October 10, 2003, 16:36:02
 Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 11:49:11 : Search time 61 seconds
(without alignments)
3184.940 Million cell updates/sec

Title: US-09-901-572a-4

Perfect score: 6413
Sequence: 1 MNISKKLKSYTLIGLAVFG.....APTKPTAPKPAAPKPAKRE 1224

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5385	84.0	1062	AAW22721	Antigenic protein
2	507.5	7.9	1122	AAW64927	Cytadhesin protein
3	337	5.3	1144	AAW6059	Mycoplasma pirum a
4	328	5.1	1627	AAW47911	Mycoplasma pneumoniae
5	324	5.1	1627	AAW47911	M. pneumoniae pl. c
6	324	5.1	1627	AAW47911	Cytadhesin pl. My
7	252	3.9	10182	ABP8314	Staphylococcus epi
8	223	3.5	2893	AAW98828	H. pylori GHPD 148
9	223	3.5	2893	AAW71556	Helicobacter polyp

10	214	3.3	2902	22	AAW46351
11	207.5	3.2	1180	23	ABW54070
12	205.5	3.2	2122	24	ABW08784
13	205.5	3.2	2344	22	AAW37120
14	205.5	3.2	10498	24	ABJ19119
15	205	3.2	2123	22	AAE00701
16	203.5	3.2	2053	22	AAW69135
17	202	3.1	2870	21	AAW95559
18	202	3.1	3178	21	AAW95556
19	199.5	3.1	1844	21	AAW18250
20	199	3.1	1612	19	AAW65088
21	198	3.1	1962	12	AAW10563
22	197.5	3.1	5795	22	AAW37017
23	197	3.1	1778	22	ABW52677
24	196.5	3.1	2150	23	AAW02566
25	196	3.1	1959	12	AAW10552
26	196	3.1	1962	12	AAW10560
27	196	3.1	1962	12	AAW10561
28	196	3.1	1962	12	AAW10557
29	196	3.1	1962	12	AAW10558
30	196	3.1	1962	12	AAW10559
31	196	3.1	1968	12	AAW10941
32	196	3.1	1974	12	AAW10940
33	193	3.0	2586	22	ABW6878
34	192.5	3.0	2261	24	ABW18914
35	192.5	3.0	2283	24	ABW6876
36	192	3.0	1536	15	AAW63505
37	192	3.0	1536	21	ABW18846
38	191	3.0	1536	18	AAW30293
39	190	3.0	2710	17	AAW5016
40	190	3.0	2710	19	AAW68387
41	189.5	3.0	1436	21	ABW18199
42	189.5	3.0	2353	17	AAW93933
43	189	2.9	1536	14	AAW41723
44	188	2.9	1072	23	ABW4963
45	188	2.9	2411	21	AAW23860

ALIGNMENTS

RESULT 1	
ID	AAW22721 standard; Protein: 1062 AA.
XX	
AC	AAW22721:
XX	
DT	27-FEB-1998 (first entry)
XX	
DE	Antigenic protein derived from Mycoplasma gallisepticum.
XX	
KW	Mycoplasma gallisepticum antigen: epitope; monoclonal antibody; bird;
KW	Mycoplasma Infection; Viral vaccine; avipoxvirus; herpesvirus.
XX	
OS	Mycoplasma gallisepticum.
XX	
PN	W09724370-A1.
XX	
PD	10-JUL-1997.
XX	
PF	27-DEC-1996; 96W0-JP03863.
XX	
PR	28-DEC-1995; 95JP-0352754.
XX	
PA	(JAPG) NIPPON ZEON KK.
XX	
PI	Fujisawa A, Yoshida S;
XX	
DR	WPI: 1997-363621/33.
XX	
DR	N-PSDB: AAT75087.
XX	
PT	Antigenic protein derived from Mycoplasma gallisepticum - useful in vaccines against, and for diagnosis of mycoplasma infection in birds

XX Claim 3: Pages 24-31; 45pp; Japanese.

CC This sequence represents an antigenic protein of Mycoplasma gallisepticum
 CC having an epitope specifically recognised by the monoclonal antibody
 CC (Mab) 35A6. This antigenic protein inhibits the metabolism of
 CC bird infecting mycoplasmas. The antigenic protein and the DNA molecule can
 CC be used in the preparation of component and viral vaccines against
 CC mycoplasma infection in birds. Mycoplasma infection can be diagnosed by
 CC hybridising DNA in a sample with a DNA probe corresponding to 10 or more
 CC (preferably 15 to 40) bases of the antigenic protein gene.

XX
 XX Sequence 1062 AA:

Query Match 84.0%; Score 5385; DB 18; Length 1062;
 Best Local Similarity 86.0%; Pred. No. 0;
 Matches 1053; Conservative 2; Mismatches 7; Indels 162; Gaps 1;

QY 1 MNISKKLKSYTLIGLAVFGALGASFGKQSDKSDNTQLYNQARTLDANSYRLAGLQ 60
 Db 1 MNISKKLKSYTLIGLAVFGALGASFGKQSDKSDNTQLYNQARTLDANSYRLAGLQ 60
 QY 61 NGSLFNTVLDVNDNFNTANGTTLKDSFTKPLGGLSDDCGGKVKQIYSDYTTSSN 120
 Db 61 NGSLFNTVLDVNDNFNTANGTTLKDSFTKPLGGLSDDCGGKVKQIYSDYTTSSN 120
 QY 121 RFDRQRTAVYALVNDENVHLKRIINTNSRIGNRNNSKFIYGVNDPAHYIRFTDGG 180
 Db 121 RFDRQRTAVYALVNDENVHLKRIINTNSRIGNRNNSKFIYGVNDPAHYIRFTDGG 180
 QY 181 TKFNTQTOGEIYNDITLAPILPKDLHPDWNTLYIQRIKLPNDVNTAVVPWVGVRVSG 240
 Db 181 TKFNTQTOGEIYNDITLAPILPKDLHPDWNTLYIQRIKLPNDVNTAVVPWVGVRVSG 240
 QY 145 ----- 144
 Db 145 ----- 144
 QY 241 TNADGMEGCGQITNTDPIAQTKTITDQNPSTFSGAMPGANNRYDSQLNVKRIKT 300
 Db 241 TNADGMEGCGQITNTDPIAQTKTITDQNPSTFSGAMPGANNRYDSQLNVKRIKT 300
 QY 145 ----- 144
 Db 145 ----- 144
 QY 301 SFQDERINTNSRIGNRNNSKFIYGVNDPAHYIRFTDGGKFNFTQTOGEIYNDFI 360
 Db 145 -----RINTNSRIGNRNNSKFIYGVNDPAHYIRFTDGGKFNFTQTOGEIYNDFI 198
 QY 361 LDAPILPKDLHPDWNTLYIQRIKLPNDVNTAVVPWVGVRVSGTNAADGMEGCGQITNT 420
 Db 199 LDAPILPKDLHPDWNTLYIQRIKLPNDVNTAVVPWVGVRVSGTNAADGMEGCGQITNT 258
 QY 421 DPIAQTKTITDQNPSTFSGAMPGANNRYDSQLNVKRIKTSQDDEKRYVPEWTSSEE 480
 Db 259 DPIAQTKTITDQNPSTFSGAMPGANNRYDSQLNVKRIKTSQDDEKRYVPEWTSSEE 318
 QY 481 NKNTIRLANSPLPSERWYIILDIPGTPVTLKEDSVNFSRLYNSVNSLSFIIDSIIYIF 540
 Db 319 NKNTIRLANSPLPSERWYIILDIPGTPVTLKEDSVNFSRLYNSVNSLSFIIDSIIYIF 378
 QY 541 GTSELPLMYYSFPTRLSDTLALNOVKTDIEASTDNGTTTNGTTTADTSSSTGAGT 600
 Db 379 GTSELPLMYYSFPTRLSDTLALNOVKTDIEASTDNGTTTNGTTTADTSSSTGAGT 438
 QY 601 GWTITTSQTVSNPTLNTYRSFGIDSKPTSAKIDETNADANVIEARIYAEYRLGIONEI 660
 Db 439 GWTITTSQTVSNPTLNTYRSFGIDSKPTSAKIDETNADANVIEARIYAEYRLGIONEI 498
 QY 661 PITNAGNFIRNTIGVGFTSGSRVYLASVNGDPTGNFOPELYVFGYGYOQTRTGT 720
 Db 499 PITNAGNFIRNTIGVGFTSGSRVYLASVNGDPTGNFOPELYVFGYGYOQTRTGT 558
 QY 721 FWYGYKLLNSPYDVLSPRYGTEINOFRRSTSLYPWAGVLTREGARSPNPFYIAQ 780
 Db 559 FWYGYKLLNSPYDVLSPRYGTEINOFRRSTSLYPWAGVLTREGARSPNPFYIAQ 618
 QY 781 GQTPESRSIFOSGYSNTEYEQSVLGEFDGIRNNLVGKASSFLNSRNPNGLEMTAA 840
 Db 619 GQTPESRSIFOSGYSNTEYEQSVLGEFDGIRNNLVGKASSFLNSRNPNGLEMTAA 678

QY 841 TTYLRSGIGLARTSGLPNQCPFGTTHOVLSVSPGDQSSIKNIRITIFPGNOLMYELETFNE 900
 Db 679 TTYLRSGIGLARTSGLPNQCPFGTTHOVLSVSPGDQSSIKNIRITIFPGNOLMYELETFNE 738
 QY 901 NKSQSVYTLRLADSSNDPDASSSPSPSLIDVNEIGVILPLDSSFTVNAAGNVALFSSN 960
 Db 739 NKSQSVYTLRLADSSNDPDASSSPSPSLIDVNEIGVILPLDSSFTVNAAGNVALFSSN 798
 QY 961 PGSPGSYTAVNTFNQNSIDIAFEGSGAKKYSIDPWGTTQFEPDEYLLONGFTSOVARNFYT 1020
 Db 799 PGSPGSYTAVNTFNQNSIDIAFEGSGAKKYSIDPWGTTQFEPDEYLLONGFTSOVARNFYT 858
 QY 1021 NOSFLNSLVDFTPPNAGCTNRYVVYDPPGNLTNQNPLKVOIOYLDGKRYDAKLKNNLVY 1080
 Db 859 NOSFLNSLVDFTPPNAGCTNRYVVYDPPGNLTNQNPLKVOIOYLDGKRYDAKLKNNLVY 918
 QY 1081 FSYNNFGALPSWVVPYTAIGSTLILAIMIILGLAIGIPLRAOKRLDQKRTFFKRYDVL 1140
 Db 919 FSYNNFGALPSWVVPYTAIGSTLILAIMIILGLAIGIPLRAOKRLDQKRTFFKRYDVL 978
 QY 1141 TAAVGSYVKRTITQTANVKKRPAALGAGSGDKKPAAPAKPAPAKSPAKPPTG 1200
 Db 979 TAAVGSYVKRTITQTANVKKRPAALGAGSGDKKPAAPAKPAPAKSPAKPPTG 1038
 QY 1201 PKSGAPRTAPKPAAPKPTAPKE 1224
 Db 1039 PKSGAPRTAPKPAAPKPTAPKQ 1062

RESULT 2
 AAR64927
 ID AAR64927 standard; Protein; 1122 AA.
 AC AAR64927;
 DT 10-AUG-1995 (first entry)
 DE Cytadhesin protein.
 DE Cytadhesin protein; Immunoassay; antigen.
 KW Mycoplasma gallisepticum.
 OS US378820-A.
 PN 03-JAN-1995.
 PD 09-NOV-1992; 92US-0973257.
 PE 09-NOV-1992; 92US-0973257.
 PR 09-NOV-1992; 92US-0973257.
 XX (DOHM/) DOHMS J E.
 XX (KEEL/) KEELER C L.
 XX DoHms JE, Keeler CL;
 XX WPI; 1995-051314/07.
 XX N-PSDB; AA081778.
 DR Nucleic acid encoding cytodhesin protein - used as a probe to
 PT diagnose Mycoplasma gallisepticum infection in poultry
 XX Claim 1; Column 7-16; 13pp; English.
 XX The protein is a cytodhesin protein from Mycoplasma gallisepticum,
 CC and may be used as a diagnostic antigen (optionally in recombinant
 CC form) in immunoassay formats for diagnosis of M. gallisepticum
 CC infection in poultry, e.g. turkey or fowl.
 XX Sequence 1122 AA:
 Query Match 7.9%; Score 507.5; DB 16; Length 1122;

Best Local Similarity 22.7%; Pred. No. 3.9e-23;
Matches 313; Conservative 172; Mismatches 472; Indels 419; Gaps 69

QY	5	KKLKSYLLIGLAEVAGLGSASFGRKQSDKSDNDNTQOLVNOARTLANSVRLAGLQNGSL	64
Db	2	KKL-TEKLSVIGITPLALIGLGSFGLANSKAPNNLKPYNQUGEMNSQ	52
QY	65	ENTVLARDVDNF-----ITANGTIIKLDSFTKPLYG-----LDLS	100
Db	53	LEKARMRNRNSFTSLSDIGTNPBALVLGKSSISRID-----LYGNVWTFDPGNNDLT	107
QY	101	DDCGY-----KVQIVSDY---TTSRNFDDQRQTRAYVALVNDKANHLKRI	146
Db	108	GRVGFYDANKKLTAFSGDVSNFSDLSKTYEATQDQEDDPVFFYLLMPDAA	160
QY	147	NTNSNRIGNRNNSKFEVIGVDNPAHVIRFDDGKRFNFNOTQGEIV-NDLILDAPILP	205
Db	161	-----VOEQTKDQVFNTRYMSDAP	181
QY	206	KDLHPDMYNLYIQRKILPNDVNTAVPWPVGRVSGTNADDMFCNGOITNTDPIAQRK	265
Db	182	-----AAGDTSABEGSNPAGGSGSSA-----AGGAVA-PAASS	216
QY	266	TTTTONQNPSTNSGAM-PGANNR-----YDSQLNKHKIKTSFOLDE-----R	307
Db	217	TARLVEENSGAGMTPTASTSETVIDYNSDQNKIPRKTKKDSSESSESIINGRTYAN	276
QY	308	INTNSNRIG-----NR-NNNSKFEVIG-----VDN-PAHVIRFDDGTFKN	346
Db	277	INTQKNQGVLYVKNENKFNSENPFAVENMAFIKKDMVDNTPSWTQOSANGKMTNVKQ	336
QY	347	FTNOTQGEIVNDFILDAPILPKDL-----HPDYNLYIQRKILPNDVN	389
Db	337	FYKHDNPAVANNRFYRAKYPRKLETOJTTPPLIDSSFPYEHPEMY-----EEN	385
QY	390	TAVVPWP--VERVSGTNADDMFC--GNGQITND--PIAQITTTDNO-----NPS	436
Db	386	QFVPMQOYIYNKGLYAKDGMVYLFEGNGTWNVNESALSIGVFTRKFNRRLEAPGNRK	445
QY	437	T--FNSGAMPANNRRYDSQLVWKHKIKTSFOLDEKFEVYPWETGSEENKINTRLATGSLPS	494
Db	446	TVGTFYGLILSAIS-FDATRNGLALAAAKQDVYHF-----VPRLAGVGS	493
QY	495	NERYWILDIPGTPOVLTKEKDSVNFSLRYLNSVLSFTGDSIYJFTSE--LPSLWY	551
Db	494	P-----RCANGNIFLGSAL-TWGTNGNFIIDKMH	522
QY	552	SFPTRLSLTALNOVKTTDDIEASST--DNGTTMGITTTADTSSGS-----TGAGTG	601
Db	523	-SPAVIED-----APTFTVYNSSGVLVNSQOOSTSPMNSNGNESIPYRWNTSYDY	575
QY	602	NTTMSQTVSNPTLWTVRSFGIDSKPTSAKIDETNMADPNYI--EARLYAEV-----	652
Db	576	NSVFPALISRPAGNTKQ--VESFETTRALKIDTLN-SLPKNFTQENNIFFSYAMLDGQ	632
QY	653	-RIGIQNE--IPTNAGNFIRNTIGVGTSTGSRVYLASYN-----DQRTGNE	701
Db	633	WSLTKRKDSTWLTNTINNFTYNTQOOLASTPAAGENANPRNLMLLTAKGDRDRIGV	692
QY	702	QPELY-----VEGVLGYOQTRGTGW---YGYIKLNSPYDVLDSFRKGTETNOQR-KT	752
Db	693	D-ILYSNNTNKFYY-YOVGALITWPEVOVNYKYSANITTYNLRTFDGSGTTPATODAN	750
QY	753	SLTFVPMGVGLTEEGARS--FSNPFYIRAOQTPESRSIFQSGVONTREYIOSVLCFDG	810
Db	751	TVSSKLNAVILASTDQOGWYNGSIYVKKASTPSSQ-----GYTMQDKGLGTTASNAV	805
QY	811	IRNNLNVGVKASSFLNSRPNPNGLEMTAATTVLYRSQIGLARTSGLPNOQPGTTHQVYS	870
Db	806	ISNNTKAGY-----SIRPDD-----TFVYSKI-----PF-----	831
QY	871	VSPPDQFSINIRIIFPGNQLWYLFITNENKSSVYTLIRLADSSNPDDASSSPISLID	930

Db	832	---	EKEITAAVNRASLDS-----	YVVLNGERTSVMDVARV---	SPDSSA-----	LT	872
QY	931	VNEIGVILPLLDNSFTYVNAAGNALFSSNDPGSGSTAAVNTFQNLSDIAFECSGAKYT	990	1	1	1	1
Db	872	LNPKRINPLMNRD----	NVIGOGAFISRN-DIPSSF-----	FENKINDIV-----			912
QY	991	SDFNGTIQPKRDEVLIONGFISQVARRFVYNQSFNLSDVETPA--	NAGTIVRYVDPDG	1048			
Db	913	-----	TTEDAGTEVLDSKINSIYRTTPQNNPHIRLRLVIDRS	952			
QY	1049	NLTQNLPLKVOIQYLDGKYVDALKNNNLVT--FSYNNFGALPSSVVPFTAIGSLILAI	1107				
Db	953	RATNDFIKLPLQV-LVDEGYAAVQANSVYFSDDEFGLFDLPGYVLPVPAISIRIIIAL	1011				
QY	1108	MIILGALIGIPLRAQRRLQDGEFTTEKKVDTLTAAGSVYKKIITQTA--NVKKKPAAL	1165				
Db	1012	ALALGILGIPMSQNRKMLKQGFYISKKKVDILTLTAAGSVFKQIINTSVINIKKTQML	1071				
QY	1166	GAGSGSGKKRPAAKKPAAPKPSAPKASSPAKPTGSPSGAAPTKEPTAP--KPAAPKPT	1220				
Db	1072	QA-----	SSPSKPSAPAKKPTGPKP--SAGCAKPTAPKAPAPPT	1118			
RESULT 3							
ID	AAR76059	standard; Protein: 1144 AA.					
XX	AAR76059:						
AC							
XX							
DT	25-MAR-2003	(updated)					
DT	08-SEP-1995	(first entry)					
XX							
DE	Mycoplasma pirum adhesin protein.						
XX							
KW	Adhesin: cytoadherence; Mycoplasma; human immunodeficiency virus;						
XX							
FT	HIV; vaccine; infection.						
XX							
OS	Mycoplasma pirum.						
XX							
FT	Key	Location/Qualifiers					
FT	Peptide	201..219					
FT	Peptide	/label= immunogenic					
FT	Peptide	1055..1072					
XX		/label= Immunogenic					
XX							
PN	FR2705970-A1.						
XX							
PD	09-DEC-1994.						
XX							
PF	04-JUN-1993;	93FR-0006745.					
XX							
PR	04-JUN-1993;	93FR-0006745.					
XX							
PA	(INSP) INST PASTEUR.						
XX							
PI	Bahraoui E, Blanchard A, Ferris S, Montagnier L;						
PI	Tham TN;						
XX							
DR	WPI: 1995-024735/04.						
XX							
XX	N-PSDB; AAO79746.						
XX							
PS	Claim 5; Fig 1; 55pp; French.						
XX							
CC							


```

FT /label= pref._fragment
XX US502636-A.
XX
XX 25-JUN-1991.
XX
XX 10-NOV-1987; 87US-0118967.
XX
XX 10-NOV-1987; 87US-0118967.
XX
XX (TEXA ) UNIV OF TEXAS BOARD.
XX
XX Baseman JB, Su CJ, Dallo SF;
XX
XX WPI: 1991-207465/28.
XX
XX N-PSDB: AAQ12518-20.
XX
XX Mycoplasma pneumoniae P1 protein and gene - used for studying
XX mycoplasma infection and for producing diagnostic reagents and
XX vaccines
XX
XX Disclosure: Fig 6(A-N); 39pp; English.
XX
XX Amino acid rpp is encoded by either TGG or TGA.
XX The peptides can be used for producing antibodies for e.g.
XX diagnosis. The MP P1 protein has a mol. wt. of 160-175 kD.
XX and has a cytohesin epitope.
XX See also AAQ12518-20.
XX
XX Sequence 1627 AA:
XX
Query Match 5.1%; Score 324; DB 12; Length 1627;
Best Local Similarity 21.0%; Pred. No. 2,4e-11; Indels 458; Gaps 68;
Matches 267; Conservative 146; Mismatches 398;
QY 251 GNGQIT---NDPI-----AQTITTDNONPSTFNSG 279
DB 506 GNGHYSKAHTAPLSIGFVRVYATGSAVTGWPYALLFSAVNKOTGLKDLDFN-- 563
QY 280 AAMPANRYDSOLNVKHKIKTSFOLDERINTNSNRGNNSKFFVIG--VDNPAHVI 336
DB 564 -----NNRW-----FEVPRMAVAAGAKFYGR---ELVLAGITTMGDATVP 601
QY 337 RFTDD--GTRKFTNOTGELYNDEIL-----DAPI-----LPKDLA- 371
DB 602 RLLYDELESNLELVAGGGLKREDQLPTPGMANRPDLIGAMSSSSSSSHNAPYEFH 661
QY 372 -PDYMYLIQRIKLPDVTAVVPPVGRVSGTNAD-----GMPGCGGQITN- 419
DB 662 NPDMDRPIQ-----NVYDAFIKPWE---DKNGKDDAKYIYPRYSGMW---AMOYIWM 709
QY 420 TDPIAQTKTITD--NONPSTFNS-----GAMP-----GANN-----RYDSQ 453
DB 710 SNKLTDOPLSADVENEMAYQPNLFAIILPELLALPDKVYKGENEFANNEYERENOK 769
QY 454 LNVK-----HRIKTSFO-----LDEKFVYEWGSEENKNIIRLATGSLPS 494
DB 770 LTVAPQTGTNMSHSPFLSRSFTGFMLVGSVLDVLYVWIGNGIRG-----N 819
QY 495 NERWILIDIGTPOVYLKEDSVVFSRLVNSVNSLSFI-----GD 535
DB 820 NHR-----GVDDITAPQTSAGSSSGISINTSGSFLPTSNIGVGLKANVQATLIGS 872
QY 536 SIYIFGTSELPLMYSFPTRLSDLTALNQVKT-----DDIEASSTDNQITTINGTITAD 590
DB 873 QTMITGGS-----PRRTLDQANL-QLMTGAGWRNDKASSGSGSDENHTKFSATGM 921
QY 591 TSSSGTAGTGNTNTTSC-----TVSNPLNTYHSRGIDSKPTSAKIDETNMAD--PNVLE 645
DB 922 DOGGSGTSGAGNPNLSKODNISKSGDSLTTQDGNALDQO-----EATNYTLPPLNP 974
QY 646 ARIYAEYRLGIGNEIPIITNAGN-----FIRNIGGVGFTSTGSRVYLRSYIGDORPTG 699

```

```

DB 975 T---ADW-----PNAISFTNNKNAQRAQLFLRGLLGI-----PYLVNRSGSDSNKRFQA 1020
QY 700 NFQPLVYVGYLYGQOGRGTGFWYGYK-ILN-----NSPFDVLDSPRG- 743
DB 1021 TDQKWSYF--DLHSQOFTKLNPAYGEVNGGLNPALVETVETGNTRAGSGSNTTSSPGICF 1078
QY 744 --TETNORRRISLYTP-----VNGGYLTEEGARSFNTPYIRAO 780
DB 1079 KIPQONNSKATLITPGIAMPDQVGNLVSGTIVSPQOLGWLVL-----FTEDVYKPR 1131
QY 781 GD-----TPESKSIQSGSDNTYEXIOSVGLGFDGI-----RNNL 815
DB 1132 AGYLGLQLGLDASDATORALIMAPRPAWAFRGSWV-NRIGRVSDVLDKGMADQASD 1190
QY 816 NVGVKASSFLNNSNRPNGL-----EMIAATYVL-RSQIGLARTSGL----- 856
DB 1191 SOGSTTATRNALPEHPNALAFQVSVVESAATKPTSSGQIQOTNSSPYLHLVKKRVQ 1250
QY 857 -----PNO-----QPFQTHQVIVSPPGQFSSIKNIRTIPEGNQWYELF 897
DB 1251 SDKLDDDLKNLDPNQRTRKLQSFQTDH---STQPOFO--SLKTTTVEFG----- 1296
QY 898 TRENKSSVYTLRLADSSNPDASSFPSLIDVNEI-----GYLLPLDINSFYVNAAGN 953
DB 1297 TSSGNLSVLS-----GGAGGSSSGSGGVDLSPVEKSGMLVGQLPST-----SDGN 1346
QY 954 VALFSSNPGSPGSYTAVNTPNO-----NLSDAFESSGAKYSDPFGTIOFKPDEYLION 1008
DB 1347 TS--STNNLAPNT---NIGNDVYVGVGRUS---ESNAKMKDDVDGIVRPLAELLIDGE 1396
QY 1009 GFTSOVARNFVT-----NOSFLNSLVD--FTPAN-----AGNRYV- 1042
DB 1397 GQTDADTGPQSVYKESPPQIDFNRLFTHPYTDLDPVTMLVYDQYIPLFDIPASVNPKNV 1456
QY 1043 ---VVDPDGNLTNOMLPLKAVQIOTLQKYYDALKN-----NLVYFS----- 1082
DB 1457 RLKVLSPDTN--EQSLGLRLEFEKRPD---ODTOPNNVOVNPNGPFLPLTASSGQPR 1511
QY 1083 -YNNFALPSSVNVPTALGSTGLIAMIILGIAIGIPLRAORKLQDKGRTTKKVDILT 1141
DB 1512 LESPONMPDVLPLAIVPLVYLVSTGLAIGIMHNRKQALKAGFALSNOKVAVLT 1571
QY 1142 AAVGSVYKRIITQGANVKKRPALGAGKSGDK--KPAAPAPAPAPKPSAPKASPPTG 1200
DB 1572 KAVGSVFEIINRT-----GISQAKRLKQTSAAKPGAPRPVPPKAPKRPVQ 1621
QY 1201 PKSGAPTKP 1209
DB 1622 P-----PKRP 1626

```

RESULT 6
 AAR67538 standard; Protein: 1627 AA.
 ID AAR67538;
 AC AAR67538;
 XX 25-MAR-2003 (updated)
 DT 07-AUG-1995 (first entry)
 DE Cytadhesin P1.
 XX P1 protein; cytoadhesin; mycoplasma; hybridization; probe.
 XX Mycobacterium pneumoniae strain M129.
 XX Key Location/Qualifiers
 FT Peptide 1..59
 FT US5369005-A. /label= sig-peptide
 PN 29-NOV-1994.
 XX

XX 22-OCT-1992: 92US-0965055.
XX
XX 09-JAN-1987: 87US-0004767.
PR 19-NOV-1987: 87US-0118967.
PR 27-JUL-1990: 90US-0558886.
PR 22-OCT-1992: 92US-0965055.
XX
XX (TEXA) UNIT TEXAS SYSTEM.
PI Baseman JB, Dallo SF, Su CJ;
XX WPI: 1995-014162/02.
XX DR N-PSDB; AAQ79074.
XX
XX Detection of mycoplasma contamination or infection - using DNA
PT encoding M. pneumoniae PI polypeptide as a hybridisation probe to
PS detect pathogenic Mycoplasmas
XX
XX Disclosure: Fig. 6A-6N; 47pp: English.
XX
XX Cytahesin P1 protein was purified from cells of virulent
CC hemadsorbing Mycoplasma pneumoniae M129. Oligonucleotide probes
CC based on the N-terminal sequence of P1 were used in Southern blots
CC to isolate P1-encoding DNA, which was then cloned and sequenced
CC (AAQ79074): the predicted amino acid sequence is given in AAR67538.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 1627 AA:

Query Match 5.1%; Score 324; DB 16; Length 1627;
Best Local Similarity 21.2%; Pred. No. 2.4e-11;
Matches 267; Conservative 148; Mismatches 403; Indels 444; Gaps 68;

QY 251 GNGQIT--NTDPIA-----QKTTTIDN-----QNPSTFNSGAMGANN 286
DB 506 GNGHVTSAHATPLSGIFRVRNATGTSATVGMPEYALLFSGMVKKQFDGLKDLDPNN 565
QY 287 RYDSQALNVKHRITTSQDERINTNSNRIGNNNNSKFTVIG--VDNPAHYRFTDD-- 341
DB 566 RM-----FEYVPRMAVAGAKFYGR--ELVLAGITTGDTATVRLYDEL 608
QY 342 GTFNFNTQGEIVNDFIL-----DAPI-----LPKDLH--PMWYNL 377
DB 609 ESNLNLVAGQGGLRBDLQFTPYGMANRPDLPGAMSSSSSSSHNAPYFHHNPMDQR 668
QY 378 YIORKILPNDVNTAVVPMYGRVSGTNADD-----GMFDCGNGQITN--TDPIAQT 426
DB 669 QIQ-----NVYDAFIKPMW-----DKNGKDDAKYIYPRYSGMW--AMQVYNNMSNKLTDQ 716
QY 427 KTTTD--NONPSTFNS-----GAMP-----GANN-----RYDSQLVNK-- 457
DB 717 PLASADVNEANAYOPNSLFAILPELLAALPDKVKYKKEEFANEXERNOKLTVAPTQ 776
QY 458 -----HRIKTSFO-----LDEKFEYPEMTGSEENKNTIRLATGSLPSENERWIL 501
DB 777 GTWMSHSPILSFSFGFNLYGSVLDQVLDYVPMIGNGYKX-----NNHR----- 822
QY 502 DITGTQVTLKEDSVAVFSLYLNVSNSLFT-----GDSIYIEGT 542
DB 823 ---GVDDITAPQTSACSSSGISSTNTSGRSFLPTFPGNIGVGLKANYQATLGSGQMTITGG 879
QY 543 SELPSLWYVSFPRRLSDLTALNOVKT-----DDIEASSYDNGTNGTTTATSSGSTG 597
DB 880 S-----PRRTIDQANL--OLMTGAGHRNDKASSGSDENHTFTSATGMDQOGQSG 928
QY 598 AGTANTTNTSQ---TVSNPLNTYRSFGIDSKPTSAKIDETNMAD--PAVIEARIYAEY 652
DB 929 TSAGNPDLSLKQDNISKSGSLTTQDGNALDQO-----EATVYTNLPVLPLPT---ADM 978
QY 653 RLGIQNEIPTNAGN-----FIRNTIGVGFTSTGSRVYLKRSYNGDQRPFGNFOPELY 706
DB 979 ---PALSTFTNNNAORALFLRGLGSI-----PVLVNRSGSDSNKFOATDQKWSY 1027

QY 707 VFGYLGVOOTRTGTFWNGTYK--LLN-----NSPYDLDSPRVG---TETNO 748
DB 1028 T--DLHSDQTKLNPAYGEVNGLLNPALVETTFGNTRAGSGGSNTTSSPEIGKRIEQQN 1085
QY 749 FRRTSLTYP-----VMGGLYTEEGARSFSTMPYIRAGD----- 782
DB 1086 DSKATLITTLGLAMTPQDVGNLYVSGTTVSQGLGWLV-----TFTDFKPRAGLYGLIQ 1138
QY 783 -----TPESRSIFQSGYSDNTYEYIOSVLGFDGI-----RNNLNGVKYS 822
DB 1139 LTGLDASDATORALIMAPRMAAFRCGSW--NRGAYRESYVDLKGVMADQASDSQSTTT 1197
QY 823 SFPLNSRPNPNCU-----EMIAATYTL--RSQIGARTSGU----- 856
DB 1198 ATRNALPEHPNALAFQSVVEASAYKPNSTSSGQTSNPPYHLVKKVYQSDKLD 1257
QY 857 -----PNO-----QPFQTHQIVSVSPQDQSSIKNIRTFIPGQNLWFLETTNNKS 904
DB 1258 LKNLDPNQVRYTKLRGSGFDH---STQPPQ--SLKTTTPVPG-----TSSGNLS 1303
QY 905 SVYTLRLADSSNPDASSSPISLIDVNEI-----GYLLPLDNSFTVNAAGVALPSSN 960
DB 1304 SVLS-----GCGAGGSGSGSGGVDSLPEKYSGLWGLQLPST-----SDGNTS--STN 1351
QY 961 PGSPGSYTAVNFTNQ-----NLSDIAFEGSGAKYTDGFWCTIQPKDDEVYLQNGFTSQVA 1015
DB 1352 NLAPNT---NTGNDVYVGRLS-----ESNAAKMNDVDGIVRPLAELLDGSGQNTADTG 1403
QY 1016 RNFVT-----NOSFLNSLVD--FTPAN-----AGTNYRV-----VVDP 1046
DB 1404 PQSVKRRKSPQIDQFNLFTHPTVDLFDPTVLMVYDQYIPLIDIPASVNNKMRKLKLSF 1463
QY 1047 DGNLTQNLNPLKVOIQIYLDGKYDYDAKLN-----NLVTS-----YNNFGA 1088
DB 1464 DTN--EQSLGLRLFEFKPD---ODTOPNNNVQVNPNGDFLALITASSQGPQTLFSPFNQ 1518
QY 1089 LPSWVPTAAGSLGILAIIMILGAIIGPLRAORLQDGKGFTEFKKVDTLTAAGSVY 1148
DB 1519 WPDYVLELATYTVIVIVISVTLGLAIPMKHKQALKGAFALSNQKDYLTAKAVSF 1578
QY 1149 KRIITQYANKKPPALGAGSGDK--KPAARAPAPAKPSAPKASSPAKPTGPKSGAPT 1207
DB 1579 KEIINR-----GISQAAKRFQTSAAKAPAPRPVPPKAPKRPVQ--PK 1624
QY 1208 KP 1209
DB 1625 KP 1626

RESULT 7
ABP38314
ID ABP38314 standard; Protein; 10182 AA.
XX
XX ABP38314;
XX
XX 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
XX DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Doucette-Stamm LA, Bush D;
 PI WPI: 2002-381255/41.
 DR N-PSDB: ABN90859.
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 PS Disclosure; SEQ ID 3159; 267pp; English.
 XX
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

Sequence 10182 AA;

3.9%; Score 252; DB 23; Length 10182;

Query Match Best Local Similarity 20.44; Pred. No. 1.1e-05; Mismatches 407; Indels 556; Gaps 73;

Matches 284; Conservative 148;

31 QSDKSNNTQLVNQAFLDANSYRLAGLNGSLFTVLDVDNFTTANGTIIKIDSF 90
 1398 QATKANSRLR-----SNINSNEKQLAERGSGYSKSLTRDDGKSLTASN----- 1443
 91 TKPLXGLDSDCGGYKVKQIYSDYTSRRRFQQRTRAYALLVDEANVHLKRIITNS 150
 1444 --PIVDLDELPEPDNGYGRQY-----SHS 1465
 151 NRIGRRNNNSKFEVIGVDNPAHVIRFT--DQTKRFNTNQ--TQGEIVNDFILDAPIPKDL 208
 1466 NIIYNEKNS--IVNGOYVEANGASAFNIDKYKKAANNGIMGYITKAQIYLAPSPKG- 1523
 209 HEDWNLKIOR-----KILPND-VNTAVVPWPVGR-----VSG-----T 241
 1524 -----YIEKLGQMLSTNNVINYFVPSDKVPSIT--VGYDHHFTYSGTEKNTI 1573
 242 NADGWFDCNGQIINTDPIAOTKTTDN-----QNPSTFNS-----GAMPGANRRYD 289
 1574 NYND--NYGLNTVASTDSATITMTRNNNELVGQAPNTVNSINKIVKATDKGNSIV 1630
 290 S-----QANVKHRIKTSFOLDERINTNSNRIGRRNNNSKFEVIGVDNPAHVIRFTDD-G 342
 1631 SFTVNIKPLNEKIRITSSS-----NOTPRISNTIONNANSTIEDQNKVKSLSMTKILG 1685
 343 TKFNTNQTQGEIVNDFILDAPIPKDLHPDWNLYIQKILPNDVNTAVVPWPVGRYSG 402
 1686 TR-NYVNESN-----NDVRSQVVS-KVNR-SG 1709
 403 TNADGWFDCNGQIINTDPIAOTKTTDN-----QNPSTFNS-----GAMPGANRRYD 452
 1710 NNAIVNTTTFSSGTNTTTPYKHAVLLEVFPTRITTVAGQO--FPGKGTSPPDFPS- 1765
 453 QANVKHRIKTSFOLDERINTNSNRIGRRNNNSKFEVIGVDNPAHVIRFTDD-G 488
 1766 -----LRTGPGPDAIIVWVNGQDINSNOIGRDLTHAEIFFDGETTPIRKDTYKL 1818
 489 TGLSPSN-----ERYW--TLDI-----PSTPO----- 508
 1819 SOSIPKQIYETTINGREFNSGDVAPGNFVQAVQIYPEHMDPFMAGSGSTPSRRNAGSFT 1878
 509 --VTL-----KEDSVNFSRLYLNS--VNSLSFT-----GDSIYFTGISELPSLWYSF 553

1879 KTVTVYQNGCTENVANLEKVPKPKVIDNSVYSKQGLNQQIILVRNVPQNAOYTLQO- 1937
 554 PTRLSDLTALNQVKTDDIEASSTDNGTGTGTTADTSSGSN-----GAGTGNNTNSQ 608
 1938 -----SNGTVIPENTTIDSNGLAIVTTOGLTPTGNTAKTS 1974
 609 TVSNPTLNTFRSFGIDSKPT-----SANKIDET-----KNADPN 642
 1975 MTNNVTYTKQNSSGSIASNTTDDISFSENSQVAVTGMQAKNDGIKIITGTYNNPNDEN 2034
 643 VLEARIVAEYRLQIENE-----IPIT-----NAGNF 668
 2035 SPISNIPASHSLTWNEEENSMKNIGTITKYTVTLTNHOGTRVDPPIITTYTAAKPP 2094
 669 IR-----NTIGVGFTSTGSRVYLASVNGDQRTGNF---QP 703
 2095 VHDQGRNLNCTDYYNIIFENNRLAGT-----ASWKNRQPDKNLQVON 2142
 704 FLYVEGYLG-----YQGTTRGTFF-----WGYTKILNNS---PY 734
 2143 LIALVYPEISPPLEVPYKVVYNPDMQPIYKIQVGDTPFKGTWAGYKHLENGEGPLI 2202
 735 D-----VLDSPPRYGTETNQ-----FRTSL---TYPVN-----GYLIEGAR-----SF 771
 2203 DCKWKFYNNQOSTGTTSDQMSLAYRTTFVKGTQDVVNPNSMGWQTSQAKETVTNAK 2262
 772 SNTPYIRAQ--GD---TPES-RSIFQSGYSDMTVEYIOSVGFQDIRNNLWGVKASFL 825
 2263 PNPPTTOSKGTGVTVPYGAVERNILISG-----TNDIQA--SADKIVLNK--GNKLTFFV 2315
 826 NSN-----RPNNGLEMIATYLRISOIGLARTSGLPNOQPFQTHOIVSYSPQDQ 876
 2316 KNDGKRWTVETGSPDINGI-----GPTNNGTAISLSR-----LAVRPD- 2354
 877 FSSIKIRITFPGNQLMYFLTEENKSSVYTLRLADSSN-----PDAS-S 921
 2355 --STEAATATGSGGTTI-----STASITELYKAPQPOVATHTYDNGTPIIDNRSN 2406
 922 SEFPTSLIDVNEIGVILPLDINSFYVNAAGNVA--LFSNPGSPGSIYAVN-----TFN 974
 2407 SLNPTKEVEIN-----YTEKLNGNEFQKSFYTKNNNGKWKINPKVNYVERN 2453
 975 QNLSDAIEFGSGAKYTSDFMGCTIOFKPDELLINGTSGQARNFVYNSQFLNSLVDFFPA 1034
 2454 QDNKGVYFSANTIKPNSOI--TTPKAGQNTEN-----TNPVIOA-----PA 2495
 1035 NAGTYRVRYVDPDG--NLFNQNLPLKVOIQYLDGKRYDAKLKNNLVTFTSNFGLAPSVY 1093
 2496 QHTLTINELVKEQGNVYNDINNNAVY-----PNKRVAIAIKQGN--ALPT-- 2539
 1094 VPTAIGSTLIGILAIM 1108
 2540 -NLAGGSTSHLPVVI 2553

RESULT 8
 AAM98828 standard; protein: 2893 AA.
 AAM98828:
 31-MAR-1999 (first entry)
 H. pylori GHPD 1484 protein.
 GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 peptic ulcer disease.
 Helicobacter pylori.
 WO9843478-A1.
 08-OCT-1998.

XX 01-APR-1998: 98WO-US06371.
 PF
 XX 29-JUL-1997: 97US-0902615.
 PR
 PR 01-APR-1997: 97US-0833457.
 PR 24-JUN-1997: 97US-0881227.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 XX
 PI Al-Garawi A, Kleanthous H, Miller C, Coenen RP, Tomb J;
 DR WPI: 1998-542293/46.
 DR N-PSDB: AAX14547.
 XX
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX
 PS Claim 8: Page 1827-1840; 2054pp: English.
 CC This sequence represents a Helicobacter pylori GHPD protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 CC
 XX
 SO Sequence 2893 AA:

Query Match 3.5%; Score 223; DB 19; Length 2893;
 Best local Similarity 20.5%; Pred. No. 0.00013;
 Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;

QY 14 GGLAVEG-----ALGSAPFGKQSDKSDNTOLVNOARTLDANSYRLAGLGGSL 64
 DB 287 GGYTFNKESATNMTAFSSGFNKGVSFN-GTFSNASTYFD-NQATFQKSSFNCGT 344
 QY 65 F-----NTYLKVDDNFITRANGTIKLSFTKPLKGLDLSDDCGYKQYVSD- 114
 DB 345 FTFNNQTNPTNNAHOPIONSSFSGNATTLK-----GFTNQQAFFNNS 387
 QY 115 ---YTSRNFQDQRFAYVALLVNDEAVHLKRTNSNRIGNRNNSKFIYGV- 167
 DB 388 NHQLTIONASFNNATNNTGKITEKDAFN---NTFTSVDTNNMS--VTGGVTLG 441
 QY 168 -----DNPAHVRFTDGTGKFNFTN-----QTGGELVDFILDAPL-LP 205
 DB 442 KNDLKGSTLDFGSSKITLA-QGTFNLTSLGSEKSVTLNLSGGITYSNLHNLGLT 500
 QY 206 KDLHPWYNLYIQRKILPNDVNTAVPWPVGRVSGTNADDMGDCNGQITNTDPAQTK 265
 DB 501 SALKTN-----ESLSNPQSFAGLMDITLYNGV-----TGQLLN-ENATSK 541
 QY 266 TTTDNQNSTFNSGAPRANRRYSOLNVKRI-KTSFQDERINTNSNRI-----GNRN 319
 DB 542 PTDSSPKSKSTNTQY-----YQCYKRGDITLYKQEFHSNIIITQALESGTYY 591
 QY 320 -----NNSKFIYGV-----VDNPAHVR-FTDDGTFNFTNQTGGELVND- 358
 DB 592 PPVINGSKFELDSASVYNADMPWYDHKYYIPKQNFTEGTYLPSVQIMSGSYTNSFK 651
 QY 359 -FLIDAPILPKDLHPW--YNLYIQRKILPNDVNTAV-----PWPVGRVSGTNADDMG 410
 DB 652 TTSANGSNLYIGYNSTWTDHNVSSGTVSFGDTSGLALNCHGCPWPYQCTGT--NGTY 709
 QY 411 DC-----GNQOITN-----TDPIAQTKTTDNQNP-----TFNSGAMP 444
 DB 710 SAHYVITANLRSGNIGTGGAANLIFNGVDSTINANATITTONAGIYSSSMFSTQSM 769
 QY 445 GANN-----RYDSOLNVKRIKTSFQDERKTFVPEWTSSENNKITRLANGSLPSNERYWI 500

DB 770 NSQNLGNLSNCKLSVYGTTFTEAKDKPIFNAGQAVFENTFN----- 814
 QY 501 LDIPGPQYTLAKDSVNVFSRLYN-----SVNSLSTIGSIYIFGSELPSLM 549
 DB 815 ---GGSYQ--FSGDSTLNFSSNNQNFNGSFEISAKNASFNNANFNNSASFNNNSATTSF 869
 QY 550 YSFEPTRLDLTALNOVKTKDIEASTDNGCTTNGTTTADTSSGSGAGCTGT----- 603
 DB 870 VGDFTNANSNL--OI-----AGNAVFGNSNGSNTANFNNTGTSVNIAGNATPDNVY 919
 QY 604 ---TNTS---QTVSNPT---LNTYRSEF-----IDSKP-----TSA 630
 DB 920 FNGPTNTSVKGOVTLNNTITLKNLNAFLSEDCGTFPNAHSVINAESITNGNITLVSSS 979
 QY 631 NKIDETNMADPNVIEARITAEYRLGIONEIPITNAGNFIRNTIGVGFSTGSRVYLRAS 690
 DB 980 KELEYNNAFSKMLMOLINOGH--GASSEKLVSAGNGYDVY-----YS 1022
 QY 691 YNGDQRPRTGNFOPFL-----YVEGYLG-----YQQTRTG--TFWYGYK- 727
 DB 1023 FNNQ---TYNFOEVSQNSISIRLGVNMYFDYDMEKSDHLTYONALGPMITMPSNYSN 1079
 QY 728 ---LLNNSPY-----DVLDSPRVGTETNOFRRTSLTPVPGCYLTEEGARSPSNPYIR 778
 DB 1080 NLGNANNTIYYDKSIDEVASGKTLPFKAEFSQT-----FTGQNSAIVGAKSIWT 1130
 QY 779 AGQDTPESNSIFQSGISDNT-----YE---YIO----- 803
 DB 1131 SLSDAPQSNITIFEG--DNKGAGSNDASGCHWNLOIGITGHEAKIYITGISGMR 1188
 QY 804 -----SVLGFDCIRN-----NLNVGKASS--FLNSRNPNGCLEMIATYVYLR 846
 DB 1189 ISSGGASLNFENLQGLILTLNATLYNRAAGTQSSNMFIS-----NSANIQONSTFYD 1242
 QY 847 QIGLARTSGLPNOQPGTTHQVIVSPGQFSSIKNIRTIPEGNQMLYFLETTNNKSS- 905
 DB 1243 D--TAQNGNPN--FSPNALNDFSNSFRGYGKTQSVFKFNKNAIISTNSTNLSSG 1297
 QY 906 VYTLR-----LADSSNDASSFSPTSLIDVNEGLV-----LPLDN 943
 DB 1298 LYOMAKSVLEDNSNLSVSG--TSSIKANALNLSOMASINASNSTLELQGLDLYVNDT 1354
 QY 944 SFY-----TYNAGANVAL--FSSNPSPGSGSYT--AVN-----TFNQMLSDIAFEGS- 985
 DB 1355 SSALNOSTITVNSNATINDYASLANSNGHLPNCAVNFNSANITTSLNSSIVYRGAV 1414
 QY 986 --GAKY-----TSDFWGTIQKRPDEYLLQNGFTSOVARNFTNOSPLNSLVDFTPANAG 1037
 DB 1415 SLGQFNLSSNSSLDFQGS-----SAITSMYAFNFDN-AFSQSPITFHOA--- 1459
 QY 1038 TNYRVVVDPDGNLTNQNLPKVOLOYLDGKYVDAKLNKNLNVFFSYNNFGAL 1089
 DB 1460 LDIKAPLUSLGGNLLNPN-----NSSVLD-----LKSOLV---FEDQGL 1496

RESULT 9
 AAW71556
 ID AAW71556 standard; Protein: 2893 AA.
 XX AAW71556;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Helicobacter polypeptide GHPD 1484.
 XX
 KW GHPD 1484; infection; therapy; diagnosis; vaccine; gastritis;
 KM ulcer.
 OS Helicobacter pylori.
 XX
 PN W09821225-A1.
 XX
 PD 22-MAY-1998.

XX 14-NOV-1997: 97WO-US21353.
XX 28-JUL-1997: 97US-0902615.
XX 14-NOV-1996: 96US-0749051.
PR 01-APR-1997: 97US-0831309.
PR 01-APR-1997: 97US-0833457.
PR 01-APR-1997: 97US-0834705.
PR 24-JUN-1997: 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (PLAC-) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Haas R, Kleantous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX MPI: 1998-297855/26.
DR N-PSDB: AAV52091.
XX
XX Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
XX
XX Claim 1; page 330-337; 362pp: English.
XX
XX This claimed Helicobacter pylori polypeptide, designated GHPO 1484,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC AAW114/4-7/1558) are claimed, as well as isolated polynucleotides
CC (see AAV5209-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transformed or transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention; therapeutic and/or
CC used to prevent or treat Helicobacter infection; therapeutically and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or Helicobacter in samples using e.g.
CC detecting the presence of Helicobacter antibodies; methods for
CC the polypeptides or monospecific antibodies; and methods for
CC purifying the polypeptides by antibody-based affinity
CC chromatography.
XX
SQ Sequence 2893 AA:
Query Match 3.5%; Score 223; DB 19; Length 2893;
Best Local Similarity 20.5%; Pred. No. 0.00013;
Matches 281; Conservative 160; Mismatches 473; Indels 458; Caps 73;
QY 14 GGLAVFG-----ALGSASFGKQSDKSNQTLVNOARTLDANSVRLAGLGGSL 64
Db 287 GGYTFNKFSATNNTAFSSGSFNFKGVSSFN-GTFSMASYTFD-NOATFQNSSFNGGT 344
QY 65 F-----NYLVRVDNFTTANGTIRKIDFTKPLXGLDSDCGYKVOYSD- 114
Db 345 FTENNOTNPTNNAQHPOIONSFSFGNATLK-----GFVNFQDAFNNS 387
QY 115 ---YTSRNFQDQRTAYALLVDEANVHLKRINTNSNRIGNNNNSKVIYGV- 167
Db 388 NHOQTIONMSFNATFNNTGKITIERKASFN---NTFENTSVDPNNNS--VTGGVTLGG 441
QY 168 -----DNPAHYIRFTDQGTKEFNFN-----OTGELVDFILDAPI--LP 205
Db 442 KNDLKGSTLDFGSSKRTTIA-OGTTNLTSLGSEKSVTLNLSGGITLNLNHLNGLT 500
QY 206 KDLHPDWYNIQRIKILPNDYNTAVVPMVGRVSGTNADGMEFGNGOITNTDPIAQT 265
Db 501 SAKTN-----ESLSNPSQFAOGLMDITTYNGV-----TGQLN-ENAAATSK 541
QY 266 TTTDNQNPSTFNSGAMPANRRYSOLANKHRI-KTSFOLDERINTNSRI-----GNRN 319
Db 542 PTDSPPSKSSTNSTQV-----YOGYKIGDTITTKLOETFFNSITLQALSGTYT 591
QY 320 -----NNSKFVIG-----VDNPAHYIR--FTDGTKEFNFTNQTGELIYND--- 358

Db 592 PEPVINGSKFDLSASNTINADMPPYDHHKYYIPKQNTFESGTYLLPSVQIMGYSTNSFKQ 651
QY 359 -FIIDAPILPRDLHPDW--YNYLYQRIKLPMDVNTAV-----PMVGRVSGTNADGME 410
Db 652 TFSANGSNLYIGYNSTWDHNVSSGTVSFGDTGSLNLCGCHGCFWPMYQCTGT--NGTY 709
QY 411 DC-----GNGQITN-----TPPIAQRTTYTDNONPS-----TFSNGAMP 444
Db 710 SAYHYITANLRSKGRIGTGAANLIFGVDSINIMANNTIQHAGAGYSSSMFTSTQSDM 769
QY 445 GANN-----RYDSOLNVKHKIKTSFOLDEKPYEPENTGSEENKNTIRLATGSLPSNERXYI 500
Db 770 NSQNLNGLNSNGKLSAYGTFTEKAKDKFTFNAGQAVFENTFN----- 814
QY 501 LDIPGTYQLTKEDSDVNSFRLYLN-----SYNSLSFTGDSIYIFGISELPSLM 549
Db 815 ---GGSYQ--FSGDSLNFSSNNQFNSGSEISAKAKASFNNANFNNSASFNNNAVTSF 869
QY 550 YSFPTRLSDLTALNOVKRTDIEASSTGNTGTTTADTSSGSGTAGTGNT----- 603
Db 870 VGDFTMANSNL-----QI-----AGNAVFGNSTGSOQTANFNMTGVSINISGNTFPDNYV 919
QY 604 ---TNTS-----QTVSNPT--LNTYRSFG-----IDSKP---TSA 630
Db 920 FNGPTNTSVKGYVTLNNTLKNLNLAPLSFGDGTTFPNAHSAVINIAESITNGPITLVSSS 979
QY 631 NKIDETWADPNVIEARIYAERLGIQNEIPTNAGNFRNTIGVGFTSGSRVYLKAS 690
Db 980 KEIPEYNAFSEKMLMOLINTYQGH--GASSEKIVSSAGNGYDVV-----YS 1022
QY 691 YNGDQRTGPNQFPL-----YVEGLG-----YQOQRTG--TFWGYTK- 727
Db 1023 FNNQ---TYNQEYFSONSISIRLGVNMFVQVDMKSDHLYQNALGFMTPNYSYNN 1079
QY 728 ---LNNSPY-----DVLDSPRYGTEPNOFRRTSLTYPMVGYLTFEBARFESNTPYIR 778
Db 1080 NLGNANNTIYKDSIDFVYASGKTLFTKAFESOT-----FTGQSAIYFGAKSIWT 1130
QY 779 AQGTPESRSIFGSGYSDNT-----YE---YIO----- 803
Db 1131 SLSDAPQSNNTIIRFG--DNKGAGSNDASGHCHNLQICGFTGHEAOKIYITGSIRESGNR 1188
QY 804 -----SVLGFDDGIRN-----NLNVGVKASS--FLNSNPNNGLEMTAATYLLRS 846
Db 1189 ISSGGGALNFGNGLOGLILTNATLYNRAAGTSSSNMFIS-----NSANTQONSFTFD 1242
QY 847 QIGLARTGDLNQQPFGTTHOVISVSPGDDESSIKNRTIFPGNOLMYFLFTNNKSS- 905
Db 1243 D--TQONGGNPN--FSTNALNLDPSNNSFPGYGTQGSVFKFAKVAISFTNSTNLSSG 1297
QY 906 VYTLR-----LADSSNPDASSFSPTSLIDVNEIGV-----LPLDN 943
Db 1298 LYOMQAKSVLEFNSMLSVSG--TSSIKAAIMLSQNASINASHSTLELQGLDNLVNDT 1354
QY 944 SFY-----TVNAGVAL--FSSNPGSGSYT--AVN-----TFNOMLSDAIEFGS- 985
Db 1355 SSLNLNOSTINVSNNATINDYASLIASNGSHLNFGAANFPMASNTTSLNNSIYFEGAV 1414
QY 986 --GAKY-----TSPFGWTIQPKPDEYLIQNGFSTQVARNFVYQSLNLSYDTPPANAG 1037
Db 1415 SLGGQTNLSNNSLSDQGS-----SAITSNTAFNFYDN-AFSQSTITFHOA--- 1459
QY 1038 TNYRVVYVDPDGNLTNQNLEPLKQVQIYLDQKKYDAKLKNNLVTFESYNFGAL 1089
Db 1460 LDIKAPLSLIGNLNPN-----NSSYLD-----LKNSQLY--FGDQGS 1496

RESULT 10
ID AAB46351 standard; protein; 2902 AA.
XX AAB46351;
AC AAB46351;

XX 05-APR-2001 (first entry)
 DT H. pylori HPN165 protein.
 DE Microbial infection; antibacterial; Helicobacter pylori infection;
 KM vaccine; screening.
 XX Helicobacter pylori.
 OS WO200073502-A2.
 PN 07-DEC-2000.
 PD 31-MAY-2000: 2000WO-EP05024.
 PF 31-MAY-1999: 99DE-1024965.
 PR 17-JUN-1999: 99DE-1027740.
 PR 21-JUL-1999: 99DE-1034029.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 PI Apfel H, Fuchs TM, Glibbs CP, Hueck CJ, Meyer TF;
 PI MPI: 2001-049948/06.
 DR N-PSDB: AAF25628.
 XX
 PT Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 PT essential genes in defective mutants -
 XX
 PS Claim 37: Figure 15; 366pp; German.
 XX
 CC This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) (which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC fragments or an inhibitor of (IIa) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process.
 CC and the identified genes can be used for screening without purification.
 XX
 SQ Sequence 2902 AA:
 Query Match 3.3%; Score 214; DB 22; Length 2902;
 Best Local Similarity 19.4%; Pred. No. 0.00046;
 Matches 277; Conservative 157; Mismatches 449; Indels 542; Gaps 70;

DB 421 DASFNNTSFNTPVDTNNMTISGCVTLGSKNDLKNCAITLDFGSSKILTL-QGTFNLTSLG 479
 188 -----QTOGEIVNDFILADILPKDLHPDWYNYLIQKILPNDVNT-----A 229
 DB 480 SEKSVTLILNSRGITVYNNHLLNHA-----NSLTNLAKTVESSSKPQSA 523
 QY 230 VVFWPYGRVSGTNAADGMDCCGQITNTDPIAQTITTDNQNPFNSGAMGANNRD 289
 DB 524 QGLMDITVNGV-----TGQLLN-ENATSKPTDSSPSKSTNSTOV----- 564
 QY 290 SOLNVKRI-KTSFOLDERINTNSNRI-----GNRN-----NSKKEVIGVD----- 330
 DB 565 --YQVGKTKIGDTIKQLETFESHSHSIITQALESTYTPPPYINSKDLASNTINDMPW 622
 QY 331 -NPAHYI---RETDDGTAFNFTNTOGEIVNDFILADILPKDLHPDWYNYLIQKILP 385
 DB 623 YNHKYYIPKSONTFESGTYYLPVQITMGSTYNSFKOTFSASNSNLVIGYNAWTDDHVS 682
 QY 386 ND-----VNTAVVPWYGRVSGTNAADGMDCC-----GNGQITN 419
 DB 683 SDTFVAFGDTSGSALNGHCGFWPYOQTGTT--NGTYSAHVYITANLRSGNRIGTGGAAN 740
 QY 420 ----TDPIAQTITTDNQNPFNSGAMGANNRDYDQ-----LNVKHKRIK-----TSF 464
 DB 741 LIFNGVDSINIAATITQHNAGAYSSSMFTSTQNMNDSQNLNLSGKLYVGTTFETNG 800
 QY 465 QUDEKTVYPEWMTGSEENKNTITRLATGSLPSNERKWLIDIPGPQVTLKEDSVVFSRLY 524
 DB 801 AKGKRIFFNAGQATFENTFN-----GGSYO--FSGDSLFPNSN----- 837
 QY 525 NSVNSLSF--IGDSIYFEGTSELPWLWYSP-----PTRLSDTALN--QVTKDIDEAS 574
 DB 838 NQPNSSGFELGANTTFNNANFNNTSFPNNSATTSPVGDITNANSIQT-----AG 891
 QY 575 STDNGTTNGTTTAD-TSSGSGTG-AGT-----GNTTNTS---QTVSNPT---LNT 617
 DB 892 NAVEGNTSNGSQTAMFNMTGVSVINAGNATFDVNVFSPNTSVKQVTLNITLKINA 951
 QY 618 YRFGIDSKPTSKANKIDE-----TNADPVITARIAYEVR-----LGIO 657
 DB 952 PLSEFGGTVFSAHVINIGEALITGNPITLVSSKAIEYNDAPSKMLQILNYQGHAS 1011
 QY 658 NEIPITNAGN-----FIRRTIG--GVGF----- 678
 DB 1012 SEKLVSAGAGYDVVYSEFNQTYNFQEVSPRSISIRRGVGMVFDYVDMESDRLYQ 1071
 QY 679 TSTGSRVYLRASYNGDQPTGNFQPLY-----VGYLGYOQTRTG---TF 721
 DB 1072 NALGFMTYMPNSYNNN---LGNLNTIYYDNSIDFYASGKTLFTRAEFSQPTTGQNSAI 1128
 QY 722 WGTGYKLNNSPYDIVDSPR-----VGETNOF--RRT 752
 DB 1129 VFQAKNIWTS---VSDAPQSNVIIRPGDNKAGSDASGCHCNLQIGTTHYEAOKI 1184
 QY 753 SLTFPYWAGYLTPEGARFSN-----TPYIRAQGDTPRESRIFQSGYSDNTYE 800
 DB 1185 YITGSIESGNRISGGGASLNFNGLOGILLTNATLYNRAAGTQSSSNF----- 1233
 QY 801 YIOSVLGFDGIRNNLVNGVAKSSFLN--SNRNP-----NGLEMIATTVLRSQIGLART 853
 DB 1234 -----VNSANLQAOQNSYFIDDTAQNCKNPNSFALMLDPSNSFRCYVG---- 1279
 QY 854 SGLPNQOFGTTHQVIVSVSPGDQSSIKNIRITFPGQOLWFLFTNNKNS--VYTLR-- 910
 DB 1280 -----OTQSVKFAVNAVIAISFTNSNLSSGLXQOAK 1311
 QY 911 --LADSNPDASSSPTSLLIDVNEIGVI-----LPLDLSFY----- 946
 DB 1312 SVLPDNLNLSVSG---TSSIKANATILSONASINASHSTLELQGLNLNDTSSILNLQ 1368
 QY 947 -TVNAGNVAL--FSSNPGSPGSYT---AVNTFNOL-----SDIAFGCS----- 985

Db 1369 SAINVSNATINDYASLASNGSHLNFNGAVNENSANITTSISSSIVFEGVSLGOFN 1428
QY 986 -GAKYISDFMGCTIOFKPDEYLIQNGFTSOVARNEFTNOSFLNVDFTPANACTNRYVY 1044
Db 1429 LSNNSLDFQGS-----SAITSNFEAFNFDN-AFSQSPTFHOA---LDIKVPL 1473
QY 1045 DPGNULNQNLEPKYQIYLDGKRYDAKLRNNNLVTFSSYNFGL 1089
Db 1474 SLGMLNLPNNSSVYL-----NLKNSQVLF--FSDQSL 1503

RESULT 11
ABBS4070
ID ABBS4070 standard; protein; 1180 AA.

AC ABBS4070;
XX
XX 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein yhpD.
XX
XX Lactococcus lactis protein yhpD.
XX
XX Biosynthesis: biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX
XX Lactis and related species -
XX
XX
XX Claim 6; SEQ ID No 772; 2504bp; French.

XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX
XX sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The
XX
XX nucleic acid sequence is useful in the detection and/or amplification of
XX
XX nucleic acid sequence, particularly to identify Lactococcus lactis or
XX
XX related species. The proteins of the invention are useful for the
XX
XX biosynthesis or biodegradation of a composition of interest. The
XX
XX invention helps research in lactic bacteria, particularly useful in the
XX
XX production of yogurt and cheese.
XX
XX Note: The sequence data for this patent is based on equivalent patent
XX
XX WO200177334 (published 18-OCT-2001) which is available in electronic
XX
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 1180 AA:

Query Match 3.2%; Score 207.5; DB 23; Length 1180;
Best Local Similarity 21.0%; Pred. No. 0.00034;
Matches 246; Conservative 132; Mismatches 445; Indels 347; Gaps 59;

QY 15 GLAVFGALGSA-----SEFGKQSDKSNNTOLVNOARFLDANSVRLA----- 56
Db 118 GLGIFSVSATSATGFLKSTIKYDSSSLDPTKSGDAIGVLTGASKAQAALANAS 177
QY 57 -----GLGONGSLFTVVL--RDYDNTTANGTITIKDSFTKFLYGLDSDCGGYKV 108
Db 178 SQATNDGLGIRG-LPPTVFLGRDLVSNLSNVSTGS-----LTGIDSATVNGW-- 224
QY 109 KQIVSDYTSRNFRDQRTFRAYVALLVNDANVHLKRINTNSNRIGNRNNNSKYVIGVD 168
Db 225 -----TDCPGVVAIRMTDSSGTLQAANPSTL---GSA 254

QY 169 NPAHVIREFTDGTGFNFNTNQTGSEIVNDFILDAPILPKDLHP---DMYNLYIQRKILPND 225
Db 255 AFSPEVNTONGTAYTATDTSVALG---VASFTTSVQEPITVW-----TPDA 301
QY 226 VNFVAVPVPVAVRGSGTNADDGMPDCNGOITNDPIAQTFTTNDNOSPSTFNGAMPAN 285
Db 302 TMTA-----STGYTST-----LTFMLIAQT-----TSGALIGASAVNGCT 338
QY 286 NRXY--DSQLNKHKRKIKTSFQDLDERLNTNSNRIGNRNNNSKFVIGGVNDNPAHVIREFTDDGT 343
Db 339 GGYTLTHTHNLQNSLTVGFV-----GGTGGVNGLSVSLNGA-----TITGS 380
QY 344 K-----NFTNQTGSE-----INDFILDAPILPKDLHPDMYNLYIQRKILP 385
Db 381 KGTPEVPIYLNKYTGQPIPSMAQSSITANNDTIGALTPASSSTSPD----- 428
QY 386 NDVNTAVVWPVGRVSGTNADGMPDCNGOITNDPIAQTFTTNDNOSPSTFNGAMP 445
Db 429 NNTYTVVVP---AAPGYQANSA-----TTTTVLAAMPASTNTSATTV--SNFTVG 475
QY 446 ANNRYSQNLNKHKRKIKTSFQDLDERLNTNSNRIGNRNNNSKFVIGGVNDNPAHVIREFTDDGT 505
Db 476 TTN--PNAINV-----STPSQOYVLFEMAKASG---TTLA--NLPNQIAYGATGT 521
QY 506 TPQVTLKEDSVNVFSLYNSVNSLSFICDSIYITGTSELPBLSWYSPFTPLSDLTALNQ 565
Db 522 ASVVT-----DNFNSNTLLS-----SYATALS-----NA 546
QY 566 VKTDDIEASSTGDCGTTTNGTTTADTSSGAGTNTNTSQTV--SNPTLNTYRSFICD 624
Db 547 VPT--GVNITQISNGTTTSGATTAALTA-----FTSQPVTGATSTLNTVY-- 591
QY 625 SKPTSAKIDETNMADPNVIEARIVAEYRGLQNEIPTNNAFNIRNTIGGVTGSGR 684
Db 592 --TWTFAVSQT-----ATYVGYTSTALNTPALPALTSTG-----TTGGS 631
QY 685 VILRASYN-GDORPTGNFOPFLVFGYLGQOYRTGTGTFWGYTKLLNSPYVLDSPRYG 743
Db 632 VTPTGTNTSPQILPSGYISAIY-----ACTSASGT-----LISSSTT 669
QY 744 TETNQFRRTSLTYVPMGVLTTEGANSFSTPIRAQDTPES--RSIFQSGYSDWYEX 801
Db 670 NGTNAAMPSCGVYAPTNQYIYQLAPMTVAVTWTFISIDPNDDPLNVAQGGFTPT-- 725
QY 802 IQSVLGEPTGRTNNLNVGKASSFLNSRPNNGHEMIATTYLRQIGLAR-TSGLPNQ 860
Db 726 VGSILG-----GSTTNYQSOYIDNMSLNGNLSTDNVTYDGAIGKIYVGYOYOE 780
QY 861 PEGTTHQVTSVSPDQFSSIKNIRTIFFPQOMLWFLFTNNNNSS--VYTLRLAD---SS 915
Db 781 --GTTTATNAPJITTYKTFESDLAHP-----YVLAGTSSNRQLVHLMYDQTTIS 933
QY 916 NPDASSSSP--TSLIDVNEIGVILLPLDNSFPTVNAAGNALFSSNPSGSAVAVNTEN 974
Db 834 NSTSQHVPAPNTPPSSEL-----TSTTDLDGADDDGL-----SY-AINGVN 875
QY 975 ONLSDIAFEGSGAKYTSDFMGCTIOFKPDEYLIQNGFTSOVARNEFTNOSFLNVDFTPAN 1034
Db 876 ---AIVNITGCGTDWGDATNSLOLSIGTYT-----ENYVALNLGVQATNML--PS 923
QY 1035 NMG---TNRYVVPDQNTNQNLEPKYQIYLDGKRYDAKLRNNNLV--TSSYNFNGA 1088
Db 924 HSGGVTSQFLQTLDAF--SLANDTVTSQTTLEVQDA---TALTNNHHIIPRAASYNPSSD 979
QY 1089 LPSWVVPYTAIGSTL-GILAIMILIGAI 1117
Db 980 LVG--ASNADGSSVENIETNPVINGDAIGV 1007

RESULT 12
ABU08784
ID ABU08784 standard; protein; 2122 AA.

XX	AB008784;
AC	
XX	28-MAY-2003 (first entry)
DT	
XX	Moraxella catarrhalis outer membrane protein, OMP106.
DE	
XX	Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
KW	vaccine.
XX	
OS	Moraxella catarrhalis.
XX	
FT	Key
FT	Location/Qualifiers
FT	Peptide
FT	1..68
FT	/label= Signal_sequence
FT	Protein
FT	69..2122
FT	/label= Mature-OMP106
FT	Region
FT	69..111
FT	/label= N-terminal sequence of mature OMP106
FT	/note= "Specifically claimed in claim 9"
FT	Misc-difference
FT	779..880
FT	/note= "Encoded by ATCTAGGC"
PN	US2002177200-A1.
XX	
PD	28-NOV-2002.
XX	
PF	20-MAR-2001; 2001US-0813214.
XX	
PR	12-NOV-1997; 97US-0968685.
XX	
PR	03-MAY-1996; 96US-0642712.
XX	
PA	(ANTE-) ANTEX BIOLOGICS INC.
PI	
PI	Tucker K, Plosila L;
XX	
DR	WPI: 2003-328486/31.
DR	N-PSDB: ABX93525.
XX	
PT	Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
PT	producing an immune response in an animal, and as ligands to detect
PT	antibodies elicited in response to Moraxella infections -
XX	
PS	Claim 33; Page 28-34; 51pp; English.
XX	
CC	The invention relates to an isolated or substantially pure outer membrane
CC	polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
CC	polypeptide and its peptide fragments are useful for producing an immune
CC	response in an animal, and as ligands to detect antibodies elicited in
CC	response to M. catarrhalis infections. The polypeptide and its peptide
CC	fragment are further useful as active ingredients in vaccines against M.
CC	catarrhalis infections. The polypeptide is useful to prepare antibodies.
CC	An antibody binding the polypeptide is useful in immunoassays to detect
CC	M. catarrhalis in biological specimens, and also in passive immunisations
CC	against M. catarrhalis infections. An antibody binding the polypeptide is
CC	also useful to facilitate isolation and purification of the polypeptide
CC	and its peptide fragment, and as probes for identifying clones in
CC	expression libraries that have inserts encoding the polypeptide and
CC	peptide fragment. An antibody binding the polypeptide is also useful to
CC	diagnose M. catarrhalis infections. The polynucleotide encoding the
CC	polypeptide is useful as a probe to identify the presence of M.
CC	catarrhalis in biological specimens by hybridisation or PCR amplification
CC	and also to detect other bacteria that might encode a polypeptide related
CC	to M. catarrhalis OMP106. The polypeptide, an antibody binding the
CC	polypeptide, its peptide fragment and the polynucleotide encoding the
CC	polypeptide are useful as reagents for clinical medical diagnosis of M.
CC	catarrhalis infections and for scientific research on the properties of
CC	pathogenicity, virulence and infectivity of M. catarrhalis, as well as
CC	host defence mechanisms. The present sequence represents the amino acid
CC	sequence of M.catarrhalis outer membrane protein (OMP)-106.
XX	
Sequence	2122 AA;

Query Match	3.28;	Score 205.5;	DB 24;	Length 2122;
Best Local Similarity	19.38;	Pred. No. 0.001;		
Matches	282;	Conservative 156;	Mismatches 502;	Indels 507; Gaps 70;
QY	24	SASFQSDSDSNNTOLVNOARTL-----DANSRLGLCON-----GSLFNTVLRD	71	
Db	482	AATG--QLKKVNOTAESALOTFYVKAYDKNGANDSKITTVGKNNKPDGYNT-----	535	
QY	72	VDDNFITAAANGTIIKLDSEFKPLYGDLSDDCGGYKQVIVSDYTSRNRDORATRAY	131	
Db	536	-----LKL-----KGGVGDVTTETNG-----PYTF	556	
QY	132	ALLVNDENAHLKINTNSNRIGRRNNNSKVIG-----GVDNPAHVR	175	
Db	557	GLANONGILTVGNSTLINDGLSVKNTNSKKQIOVADGITTFYDINSKPGAGIEN---TTR	613	
QY	176	FTDGTGKFNFNQOGELVNDFFILDAPILPKDHPWYNLYIQRKILPNDVNTAVPWPV	235	
Db	614	ITRGC--IGFANNNGS-----LDAN-----KRLITPTGIN-----A	642	
QY	236	GRVSGTNADDMFDCGNG-----QITNDEIACQKTTTD-----NONSTFSGAMP	282	
Db	643	GCKELTVNQSLINPATNGGLDFMNRRLSTATEKSGASATIKDLYNLSQVLFTEFGDGP	702	
QY	283	GANNRY-----DSOLNVKHKRITSFOLDERINTNSR	314	
Db	703	NVTKLEIELKVKGKTTADDLTKNNIGVADSTDSNLTV-LAKTILSDLD-AVNTKTLT	760	
QY	315	IGNR-----NNSKFPVIG-----GVDNPAHVRFPD-----DG	342	
Db	761	ASDVAVYDSDGNNTAKLONGLTFESKONTGATPATNSKITIGDG-----LKFDDNNGIALDG	816	
QY	343	TKFNFTQTOGEIYNDFILD--APILPKDHPWYNLYI--QRKILPNDVT---AVPEWP	396	
Db	817	TTY--ITDKVGFAGQDSDLSKSKRYLDM-----KLKVGVEIITNINMGAKRITGLS	869	
QY	397	VGRVSGTNADDG-----MFCGNGQITNTD--PIAQTKTTDNO	433	
Db	870	NLTLDATNATLTGHTOLGIVDSTDKTRAASIGDVLNAGFNLKNNNDAKDFVSTDYDTV--	927	
QY	434	NPSTFNSGAMPGANNRROSOLNVKHKRITSFOLDERFVYPEWGTSGEENKNTIRLATGSLP	493	
Db	928	---FINGNATTAAYTYIDGKAS---KVAIDVANDGTTIHLTGADGKNKNOIGAVTTTTLT	978	
QY	494	SNERYWIIDIDGTFQVTLK---EDSVAVFSERLVLSVNSLSFIDSIYIEFTSELPSLM	549	
Db	979	KT-----DAAGDKAAINFSVNSGDDKALLNKDLADNLTLA--GEINTNGTAD-----	1022	
QY	550	YSEPTRLSDLTALNOYKTDIEIASSTDNGTGTNGTTTADTSSGSTAGTCNTNTSQT	609	
Db	1026	-----TALQTFQKKRYK---ENGDDDN---DADPTT-----VGKDAKTDNO-	1058	
QY	610	VSNPTLNTNRSEFGIDSKRPSANKIDETWMADPNVIEARIYAEYRLGIONEIPITNAQNF	669	
Db	1060	-----VNTLKLKGNKGDLDIQTNK-----DGVTTFGINTQSGLAKGNNTTLNNGLSI	1106	
QY	670	RNTIG-----GVCFSTGSRVYLRSYNGDQRPNGNFOPEFLYVGYLGQOQTPRG	720	
Db	1107	KNTAGNEIOYGADGVCAKAYNNGV--GAGIDGTRITRLRDELGFAGCTNGSLDKSRPHLSK	1166	
QY	721	--FWYGYKLLNNSPFDVLDSPRVGTEETNOFRRTSLTYPVMGYV-----TE-----	768	
Db	1166	DGINAGGKKTIN-----IQSGEIAQNSND-----AVNGKLYDKLELTKLENKISSTA	1211	
QY	769	RSEFSTPIITRAGDTPERSLIFQSGYSNTYEXIOSYLG---DQIRNNLN-----VGK	820	
Db	1212	KTAQNSLHEFVADEQGNFTVSNPSYSDTFSKTSDVITTFAGEENITTKVNGGVVRVID	1271	
QY	821	ASSFLNSNR---PDPNGLEMIATATYRLSGIARLTSGLPMQOQPGTTHOVIVSPGQOF	877	
Db	1272	OTKGLTTPKLVGNNNGKIY-----IDSQNGQNTITLSN-----	1307	
QY	878	SSIKNRTIPFGNOLWFLFTNENKSSVYTLR---LADSSNDPASSFSPTSLIDVNE	933	

```

DB      1308  -----TLANV-----TNDKGSVRTTEGKLIKEDDKTRAA-----SIVDVLS 1344
OY      934  IGVIL-----PLID--NSFYVMAAGNALPSSNGSGSYTAVNEFNONLSIDAFEGSGA 987
DB      1345  ACFNLQNGEAVDEFTVDVNFADGNATTKVYTDOTSKTSKYVYVNVDDTTIEVKOK 1404
OY      988  K-----YTSDFMGTIQFKPDEXLIQNGFT-----SOVARNEFTVNOSEFLNSVDFTP 1033
DB      1405  KIGVMTTLTSTGTGA-----NKFALSNQATGDALVYKASDIVAHMLNTLSGDITQAKGASQ 1459
OY      1034  ANAGNIRVYVDPDGNLTONLPKLVQIQYIDGKYDAK-----LKNMLVTFEY 1083
DB      1460  ANSSAGY---VDADG-----KVYDSTDNKTYYQAKNGCYDTKREYAKDKLVAQAO 1508
OY      1084  NNGALPSWVVPALIS-----NLGLTAL 1107
DB      1509  TPDTGLAQMNVKSVIKKEGVNDANKKGINEDNAFVKGLEKAAASDNKTKNAAVTGD LNA 1568
OY      1108  MTLGLALGILPRAQKRLQ---DKGKTFEKKV-DTLTAAVGSYVKKIITO----- 1154
DB      1569  V-----AQTPLTFAGDTG--TTAKKLGELTITKGGQDTNKNLNNIGVAGT 1614
OY      1155  ---TANVKKRPAALGAKSG---DKRPAAAKPAAPAKSPKASPAKPTGPK-----S 1203
DB      1615  DGFYVLAQDLTNLNSVNAAGTKRIDKGVSFYDSSGOAKAMPVLSANGLDLGKVISNV 1674
OY      1204  GAPPKPT 1210
DB      1675  GKTKPT 1681

RESULT 13
AAU37120
ID  AAU37120 standard; Protein: 2344 AA.
XX
AC  AAU37120;
XX
DE  14-FEB-2002 (first entry)
XX
KW  Staphylococcus aureus cellular proliferation protein #1290.
XX
KW  Antisense: prokaryotic cellular proliferation protein;
XX
KW  antibiotic; antibacterial; drug design.
XX
OS  Staphylococcus aureus.
XX
PN  WO200170955-A2.
XX
PD  27-SEP-2001.
XX
PE  21-MAR-2001: 2001WO-US09180.
XX
PR  21-MAR-2000: 2000US-191078P.
XX
PR  23-MAY-2000: 2000US-206848P.
XX
PR  26-MAY-2000: 2000US-207727P.
XX
PR  23-OCT-2000: 2000US-242578P.
XX
PR  27-NOV-2000: 2000US-253625P.
XX
PR  22-DEC-2000: 2000US-257931P.
XX
PR  16-FEB-2001: 2001US-269308P.
XX
PA  (ELIT-) ELITRA PHARM INC.
XX
PI  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX
PI  Yamamoto RT, Xu HH;
XX
DR  WPI: 2001-611495/70.
XX
DR  N-PSDB: AAS54979.
XX
PT  New polynucleotides for the identification and development of
XX
PT  antibiotics, comprise sequences of antisense nucleic acids -
XX
PS  Example 3; Seq ID No 12713; 511pp; English.

```

```

XX      The invention relates to antisense inhibitors of genes essential to
CC      prokaryotic cellular proliferation, their use in identifying the
CC      genes, their use in the discovery of novel antibiotics, the essential
CC      genes themselves and the encoded proteins. The prokaryotes used are
CC      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC      invention is also useful for the identification of potential new targets
CC      for antibiotic development. The antisense nucleic acids can also be used
CC      to identify proteins used in proliferation, to express these proteins,
CC      and to obtain antibodies capable of binding to the expressed proteins.
CC      The proteins can be used to screen compounds in rational drug discovery
CC      for homologous nucleic acids which are required for cell proliferation in
CC      essential prokaryotic cellular proliferation protein.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 2344 AA;
XX
Query Match 3.2%; Score 205.5; DB 22; Length 2344;
Best Local Similarity 19.2%; Pred. No. 0.0012;
Matches 226; Conservative 167; Mismatches 483; Indels 301; Gaps 50;

2 NISKKIKSYTL-----IGGLAVFAGALGSASFQKOSDK-----SNDNTQLYN 43
58 SISKMMGYGLKTTAVTGAFYVNLMLDQO-AFAASDAPLTSSELNTOSEYVGNOSTTID 116
44 -QARTLDANGVRLAGLQNGSLFNTVLRVDNDFITANGTILIKLDSFTKPLYGLDLSDD 102
117 ASSTADSTSVTKNSSSVQTSNDIYVSEKSEVISTSTNSQOEKLTLS----- 166
103 CGGYKVKQIAYSDYTSNRREDOROTRAYALLVNDPANHAKRITNSNRIGNRNNNSKF 162
167 -----TSESTSKNTSSSDTKSVSTSTSEDP-----INTSTQSTASNNTSOS 211
163 VIGGVDPNAVHIRTDDGTGFNFNTQOGIEVNDFLIDAPILPKDLHPWNYLYQKRL 222
212 T-----TPPSANLNTSTTSTS-----TAVYKLRTPS-----RLAMSTF 245
223 PNDVTAVPMPVGRVSGTNAADGMFDCGNGOI-TNTDPLAQYKKTNDONSTFNSGAM 281
246 ASAAATTALTANTITVKNKDNLKQYMTSGNATYDOSGVVTLTQDTTSOKGALTITG 302
282 PGANNRYSQILNVKHKRITSFQDLERINTNSNRIGNRNNNSKFEVIGVNDPAHVIRFTDD 341
303 -----RIDS-----NKSFFHSGKYNLGGKHYEGHGN-----GG-----D 330
342 GTFNFNTQOGE-----YNDFLIDAPILPKDLHPWNYLYQKRLPNDVN 389
331 GIGFAFSPQVGLGTCALGAAVIGLISNAFPKLDYHNSTP---NSSKAKADPSNV- 386
390 TAVVMPVGRVSGTNAADGMFDCGNGOITNTPD-IAQTKTTTND-ONPSTFNSGAMGA 446
387 -----AGGGAFF-----GAVTTDSIGVASTVSSSTADNAKLN--VQPTN 425
447 NNRYSQILNVKHKRITSFQDLERINTNSNRIGNRNNNSKFEVIGVNDPAHVIRFTDD 484
426 NTFQDFDIN-----YNGDTKVMYVYAGQVTRNLSIDIAKSGTTNLSLSMTASTGG 477
485 -TRLA-----TGSLSNERNRYMILDI-PGTPQVTLKEDSVNVFSLYL-NSVNSLSF 532
478 ATMLQOVQGTETFEYTESAVTQVRY-VVYTTKDKDILPKRTYSGNVDDQVVTIDNOQSALTA 535
533 IG-----DSIY-----IFGTSLSLWYSFPTRLSDLTALNOV-----KT-DDI 571
536 KGYNTSYVDSVASTYNDNKTVMKNAGOSYTYFTDVKAPTVVGVQGTLEVGKTNMPI 595
572 EASSTDNGTTTNGTTTATDTSSGSGTGAGGNTNTNSQVSNPTLNTYRSFGIDSKPTSAN 631
:::||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
Db 596 VLTTDNGTGTNTVTGLPSGLSYDSANSLIGPPTKIGQSTVTVSTDOANNKSTTF 655
Qy 632 KIDETNMADPNYI-----EARLYAEXRLQIOMEIPIT-----NAGNFIRNTI-----G 674
Db 656 TTNVVDTPATPTYPPIQDKSEVFS-----PLSPINIAIQDNSGNAVNTVTYGLS 705
Qy 675 GVGFTSTGSRVYLARASVNGDQRPFGNPFELVYFGYQOQRTGTFWYGYKLLNNSPY 734
Db 706 GLTFDSTNNTI-----SGTPNTIGSTITITVSTDASGNKRTTF---KKEVTNNSMS 754
Qy 735 DVLDSFRVGTETNOFRT-----SLTYVMGCYLTLEAGARFSNTPYRAQDTPESNS 788
Db 755 DSV-STSGSTQOSQSVSTKADQSQASSTSGIMTSTA-STKSTSVSLSDSVASAKS 812
Qy 789 IFOSGYSDMTYEXIGSVLGFEDGIRNLNLYGVKASSFLNSRPNPNCLE-----M 837
Db 813 LSTSESNVSSTSTSLVNSQSVSSMSGSKSTSLDPIFINSSTESKESVSTSTSDS 872
Qy 838 IATYTLRSQIGLARTSGLPNOQPGT-----THQVLSVSPGDPFSSIKNI----- 883
Db 873 LRTSTSLDSVSMSTSGSLSKSQSLSTSDASSTQSVSDSTNSISTSELSSESGSNTS 932
Qy 884 RTTFPGNQLMYFLTFEENKSSVYTLRLADSSNPDASSSEPSLIDVNEIGVILPLDN 943
Db 933 ESISISNSI-----SNSVSASTSKL-ESOSTSISLSTDSKSMSTSE-----SLSDS 978
Qy 944 SFYTVAGNVALFSSNPGSPGYTAVNPFNOMLSDIAFEGSGAKTSDPWCITQPKPE 1003
Db 979 TSTSDSVSGSLSVAGSQSVSTSTSDSMST-SEMISD-SMSTSGSLASDS-KSMVSSSM 1035
Qy 1004 YLIQNGFTSQ-VARNFVTNQSFNLSDVFTPANAGTN 1039
Db 1036 STSGSGSTSESLSDSISTSDSKSLSTSGSGSNTS 1072

RESULT 14
ABJ19119
ID ABJ19119 standard; Protein; 10498 AA.
XX
AC ABJ19119;
XX
DT 06-MAR-2003 (first entry)
XX
DE Pathogen specific antigen related staphylococcal protein SFQ ID No 440.
XX
KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW autoimmune disease; HIV; hepatitis.
XX
OS Staphylococcus sp.
XX
PN WO200259148-A2.
XX
PD 01-AUG-2002.
XX
PF 21-JAN-2002; 2002MO-EP00546.
XX
PR 26-JAN-2001; 2001AT-0000130.
XX
PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI Meinke A, Nagy E, Von Absen U, Klade C, Henics T, Zauner W;
PI Minn DB, Vytvlytska O, Etz H, Dryla A, Welchart T, Hainer M;
PI Tempelmeier B;
XX
DR MPI; 2003-075410/07.
XX
PT Identifying, isolating and producing hyperimmune serum-reactive
PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation
XX
```

```
PS Example 7; Page 234-235; 252pp; English.
XX
CC The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This sequence represents a staphylococcal protein
CC relating to the method for identifying and producing pathogen specific
CC antigens of the invention.
XX
SQ Sequence 10498 AA;
Query Match 3.2%; Score 205.5; DB 24; Length 10498;
Best Local Similarity 19.3%; Pred. No. 0.0098;
Matches 263; Conservative 152; Mismatches 507; Indels 441; Gaps 65;
Qy 40 QLVNQAARTLDANSVRLAGLQNGSLFNTYLRVDVDFNTAANGTITIKLDSFTK----- 92
Db 1104 QTVNAKATVAVTIAK-----GQSFS-----IGDIKYGF-TLSNGQPIPSGFTITSDDRTI 1153
Qy 93 -----PLYGLDSD-----DCGYYKKQIYSDYTTSRNPFDOQRATAYVALY 135
Db 1154 PTAQEVVSQMNAGTQLVHITATNAVYHKDSEDFYLSLTIID--VKQPECDQVRYTSRYDLP 1211
Qy 136 NDEANVHLKRIINTNSRIGRNNNSKFVIGGVNPNPAHVIREFTDGRFENFTNOGIEYN 195
Db 1212 TDEIS-KVQAFINANRDVYTLAEGDISVYNTENGANVSTIYNINKGRITKFSANLIAN 1270
Qy 196 DFILDAPILEPKDLHPMYNLVIQK-----LIPNDVNT- 228
Db 1271 MNFLRWVNPFDYTVVMTNAKIANRPDGLSMDHKLILRYDAFLGTQITNDILTM 1330
Qy 229 --AVVFWPVGKRGVSTNADGMEPCGNGQITNTPIAQTKTTTNOANPSTNSGAMP----- 282
Db 1331 LKATTTVPGLRNMTGSEKSAQEAAGRPNFRRTGYSSNATYTGQROFLNLGVOYLDI 1390
Qy 283 -GANNRVDSQLNVKHKRIKTSFQDERINTNSNRIGRNNNSKFVIGGVNPA----- 333
Db 1391 INPSNGYGGQ-----PVYNSN--TRANHSNSTVNVNPNPANGAGAF 1431
Qy 334 --HVIREFTDGTKEFNTQGEIVNDFILADLPKDL-----HPDWYNLVIQK 381
Db 1432 IDHVV-----KSNSTHNASDAVYKAQVLTFYGPROYEHLNONTGNTDAINIV--- 1481
Qy 382 KIILPND-VNTAVVPMWYGRVSGTNADGMEPCGNGQITNTPIAQTKTTTND---QNPST 437
Db 1482 -FVPSDLVNPETI--SVGNVT-----NHQVSGETFTNTITANDNGVQSVTV 1525
Qy 438 FNSGAMPKANRRYDSQLNVKHKRIKTSFQDEKFEVYPEMTGSEENKNTITRATSLPSNER 497
Db 1526 PNTSOTITGYDN-----NHQVSAATA-----PWT-SANNKRTINLAT----- 1562
Qy 498 YWTLIDIPGPQVTLKEDSVNVFSRLYLNSVNSLSFGDSIYIFGTSPLSLWYSEPFRL 557
Db 1563 ----DTSGNTATP-----SFNV-----TYKPLR--DKYRVGTSS----- 1590
Qy 558 SDLTALNQVKTDDIEASST-----DNGTITNGTITTDATSGSGTGAAGNTTNSQVSNP 613
Db 1591 --TAANPVRIANISNATVVSADQFTIINSLTFETVPRNSYARASANEI-TSKIVSN- 1645
```


QY 614 TLNTRSGIDSKPTSAKIDETWMDPNVTEARVTAERLIGIONEIPITNACNTRNTI 673
XX
DB 1646 -----VSRGT-----NANANTVTVYTGDSSTETVAP-----VKHVI 1677
QY 674 GGVGFTSGSRVYLASY--NGDORPGNRPFLVYVGYLGYOQTRGTFWYGYTKLNN 731
XX
DB 1678 -----PEIVASHYTVQGDFFPAGN-----GSSASDFKLSNG 1710
QY 732 SPYVLDSPRVGTETNOFRRTSLAYPVMGYLEEGAR--SESNPYTRAQGDY----- 783
XX
DB 1711 S-DLADA-----TIW-VSGQAPKNDNFRIGEDIVTAHLIDGETTPITK 1754
QY 784 -----PESRITQSGYSD--NTVEIOSY-----LGFD----- 809
XX
DB 1755 TATYKRVATPKHVEETARGVLYPVSMDYDKQVYKPVNNNSMSTNACHMFOVGTGYP 1814
QY 810 -----GIRNNL-----NVKASFLNSNRPNGLEMAIATYLSQIGLARTSGLPN 858
XX
DB 1815 NKDVVGISTRLIRVYTDNRQEDLILSKVKPDPRIANSVT-----YKAGLITN 1864
QY 859 QQ-----PQCTHOYISVSPGDOSSIKNRTIPPQ-----NQ 891
XX
DB 1865 QEIKVNNVNNSSVKLEKADNTPLVNTITHSGFSSVYTVSDALPENGIRAKSSISMNN 1924
QY 892 LMYPLFTNNKSSVYTLRLADSSNDPSSSPTSILDVNEIGVILPLDINSFTVAA 951
XX
DB 1925 VTY---TTQDEHGQVYVTRNESVDSNDATVYVTPOLATTEGANFIKGGDF---DF 1977
QY 952 GNVALFSSNP-----GSPGSY--TAVNTPNOMLSDIAREGSAKITSDFWGTIOFK 1000
XX
DB 1978 GHVERFIQNPFGATVAMHDSPTMKNTVNTKRTAVVILP--NGQTR-----NVEVP 2029
QY 1001 PDEVLIQNGFTSQARNEVTNOSFLNSLVDFTPANAGT--NRYVVDPDGNL----- 1050
XX
DB 2030 VKYVYVAN--AKAPSRVKGQNLTN-----GTDAMN--ITFDPNTNNGITAM 2075
QY 1051 TNQNLPLKQV--IQYL--DCKYVAKIKNNNLVTFSYNFGALPBMWVPTAIGSTLGILA 1106
XX
DB 2076 ANROOPNQOAGVOHLNDVDTYPCGISAKKRYPVANYQF--EPPQTYTTTGGT----- 2129
QY 1107 IMILGLAIGPLRAORLQD-----KGEKTFEKVYDILF--AAVSYVKKIITQYAN 1158
XX
DB 2130 -----LASQOASGYAHMONATGLPTDGFYKNNRPDTGTNDANMSAKPNVAKVYNA 2183
QY 1159 KKKPALGAGKSGDKKPAAPAKAPAPKAPKASSPAKPTGP 1201
XX
DB 2184 KYDVIYNGHTFATSLPAKFVYKDVOPAKPYTETRAAGAITAP 2226

RESULT 15
AAE00701
ID AAE00701 standard; protein; 2123 AA.
XX
AC AAE00701;
XX
DT 02-JUL-2001 (first entry)
XX
DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
XX
KM Outer membrane protein-106; OMP106; haemagglutination; vaccine;
XX
KW bacterial infection; immunogen; cytotoxic; antidiabetic;
XX
KM passive immunisation.
XX
OS Moraxella catarrhalis.
XX
PN US6214981-B1.
XX
PD 10-APR-2001.
XX
PF 12-NOV-1997; 97US-0968685.
XX
PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTENX BIOLOGICS INC.
PI Tucker K, Plosilla L, Tillman UF;
XX
XX WPI: 2001-281002/29.
DR N-PSDB; AAD04029.
XX
XX Novel nucleotide sequences encoding Moraxella catarrhalis outer
PI membrane protein-106 polypeptide, useful for diagnosis of bacterial
PT infections and as vaccine against Moraxella catarrhalis infection of
PT mammals -
XX
XX Claim 7; Column 53-64; 49pp; English.
XX
XX The present sequence is haemagglutinating Moraxella catarrhalis outer
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic
CC and prophylactic vaccine against M. catarrhalis infections of mammals.
CC It is used for diagnosis of bacterial infections and as reagents for
CC clinical or medical diagnosis of M. catarrhalis infections and for
CC scientific research on the properties of pathogenicity, virulence and
CC infectivity of M. catarrhalis. It is also used as a probe to identify
CC the presence of M. catarrhalis in biological specimens and to identify
CC other bacteria that encode a polypeptide related to M. catarrhalis
CC OMP106. OMP106-derived polypeptides are used as ligands to detect
CC antibodies elicited in response to M. catarrhalis infections and also
CC as immunogens for inducing M. catarrhalis-specific antibodies which are
CC useful in immunoassays to detect M. catarrhalis in biological specimens.
CC Cytotoxic antibodies are useful in passive immunisations against
CC M. catarrhalis.
XX
XX Sequence 2123 AA:
SQ
Query Match 3.2%; Score 205; DB 22; Length 2123;
Best Local Similarity 19.8%; Pred No. 0.0011;
Matches 286; Conservative 157; Mismatches 502; Indels 498; Gaps 72;
QY 24 SASPGKQSDKSDNTOLVQARTL-----DANSVALAGLQGN-----SLFNTVLRD 71
DB 482 AATYG--QLKKVQNTASALQTFYKVKVKNQNGNDANDSITIVGKNNKPDGQVNT----- 535
QY 72 VDDNFTTANGTIKIIDSFTKPLYLGLSDPGGVKQIVSDYTTSRNRPDQROTAYV 131
DB 536 -----LKL-----KGENGVDTVETNG-----TYVE 556
QY 132 ALLVNDENAMVHLKIKINTNSRIGNRNNKSFYVG-----GYDNPAAHYR 175
DB 557 GLNQNGGLFVGNSTLNDGLSVKNTNSKQIQVGADGISTFTDISNKPAGIEN---TTR 613
QY 176 FTDDGTRKNTNNOGELVNDFLDA--PILPKDLHPWYNYLQIKLIPNDVNTAYVP 232
DB 614 ITRDG--IGFANNIGS-----LDANKPRLTPFGINAGKEL-----TNVQSAINP 656
QY 233 WPGVHVSCTNDDGMEFGCGNQIT--NDPIAQOTTTD-----NONPSTENGAMPANNR 287
DB 657 -----ATNG--GOLDFMNRILSTANTEKSGSAATIDLYNLSOVPLTFPGDGPVNTYK 707
QY 288 Y-----DSQLNVKHKIKTSFQDLERINTNSNRIGNR- 318
DB 708 LGEILIKVKGCTTADDLIKNNIGVADVSTDSLVK--LAKTSLDLD--AVNTKTLTASDKV 765
QY 319 -----NNNSKF-----VIGGVNPAHYIRFTD-----DGTFN 346
DB 766 TVDSGNNTFAKLQNGDLTFGSKONTGATPATNSKIYGVG---LKFDDNNGIALDGTTY- 820
QY 347 FTNQTQGEIYNDPILD--APILPKDLHPWYNYL--ORKILPNDVNT--AVWPVYGRV 400
DB 821 ITRDKVGFAPKQDSDLSKSKPYLDKQ-----KLKVGVEITTINGINAGKAKITGASNTLT 874
QY 401 SGTNADG-----MFDGCGNQITND--PIAQTKTTTQNONPST 437
DB 875 DATNATGHTQLGIVDSTDKTRAASIGVLNAGFNLKNNGAKDFVSYDTVD----- 928
QY 438 FNSGAMPANNRDSQNLNVKHKIKTSFQDLERKVVYEWGTSSENNRITRLATGSLPSNER 497

Db 929 FINGNATTAQVYDCKAS-----KVAYDVNDGTTIHLTGADGNKNGIGKTTTLTKT-- 981
QY 498 YWIDLPQTPQYTLK-----EDSVNFSRLXNSVNSLSFGDSIYIFGISELPSLWYYSF 553
Db 982 ---DAKGKAINFESVNSGDDKALINAKDIADNLTLA--GEIRNRTGTAD----- 1026
QY 554 PTRLSDTLALNGYKTDIDIEASTDNGCTTNGTTTADTSSGSTAGTGNTNTSOTVSNP 613
Db 1027 -----TALQTFQYKVKV-----ENGDDN-----DADTIT-----VGKADAKTNO----- 1060
QY 614 TLTNYSFGIDSKPTSAKIDETNADPNVIEARIVAEYRLGIONEIPITNAGNFIKNTI 673
Db 1061 -VNTLKLKGNGLDIOGTNK-----DGVTFGINTQSGSLKAGNNTTLNNGLSIKNTA 1111
QY 674 G-----GVGFTSGRSVILRASYNDQDPTGNFOPFLYFGYLGIOQTRGT--FW 722
Db 1112 GNEQIQVGADGVFAKVNNGVY--GAGIDGTRITRDEIGFAGTNGSLDKSKPHLSKDGIN 1170
QY 723 YGTYKLNNSPYVDLDSPRVGETETNOFRRTSLTPVMGGL-----TE-----EGARSF 772
Db 1171 AGGKKTITN-----IOSGELAQNSND-----AVTGKTIYDLKTELENKISSTAKTAQ 1216
QY 773 NTPYIRAGDTPESRSIFQSGYSNDNTYEQIYSLGF--DGIRNNLN-----VGKASSF 824
Db 1217 NSLHESVADDEQGNFTVSNPSSYDTSKTSVDITFAGENGITTKVKNGVVRVGIQDTKG 1276
QY 825 LNSNR---PNPNGLMIATTYILRSQIGLARTSGLPNOQPFTHHOVISVSPGDQFSIK 881
Db 1277 LTTPKLTVGNNGKIV-----IDSONGONTITGLSN-----TLA 1311
QY 882 NIRTIFPNQWLWYFLFTENENKSSVYTLR---LADSNPDASSFSPTSILIDVNEIGV 937
Db 1312 NV-----INDKGSVRTTEOGKIIRKDEKTRAA-----SIVDYLAGEFN 1349
QY 938 L---PLLD--NSFYTVNAGVNALEFSSNPGSSTYAVTFENQNLSDIAFEGSGAK--- 988
Db 1350 LOGNGEAVDVFSTYDVFNEFADGNATYAKTYDDTSKTSKVYDVNVDDTIEVKDKKLV 1409
QY 989 ---YTSDFWGTIQFPEDEYLLONGFT-----SOVARNEVTNQSFLNSLVDFTPANAG 1037
Db 1410 KTTTLTSTGTGA-----NKFALSNQATGDALYKASDIVAHLNTLSGDIQYAKGASQANS 1464
QY 1038 TNYRVVDDPDGNLTNQLPLKVOIQYLDGKYYDAK-----LKNNNLVTFESYNNFG 1087
Db 1465 AGV---VDADGN-----KVIYDSTDNKIYQAKNDGYDKTEYAKDKLVAQAOTPDG 1513
QY 1088 ALPSWVVPYPAIGS-----TGILAIMIIL 1111
Db 1514 TLAQMNKSVINKEQVNDANKKQINEDNAFVKGLEKASDNKTKNAAYTVGDLMNV--- 1570
QY 1112 GLAIGLPLAQRKIQ---DKGFYTFPKV--DTLTAVGSYKKIITQ-----T 1155
Db 1571 -----AQTPPLTFAGDTG--TTAKKLGELTLIKGGQTDINKLTDNNIGNVAGTDFGF 1619
QY 1156 ANYKKKPAALGAGKSG---DKKPAAPAPAPAKPSAPKASSPAKPTGPK---SGAPT 1207
Db 1620 VKIAKDLTNLNSVNAAGTKIDKGVSVBSSGQAKANTPYLSANGLDLGKVISNWGKGT 1679
QY 1208 KPT 1210
Db 1680 KDT 1682

Search completed: October 10, 2003, 16:33:25
Job time : 73 secs

THIS PAGE BLANK (USPTO)


```

Db      181 TKFNFTNOTGGEIYVNDFTLDPAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSG 240
OY      241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
OY      241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      361 LDAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSGTNADGMDGCGNGOITNT 420
OY      361 LDAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSGTNADGMDGCGNGOITNT 420
OY      421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERKFEVPEWTGSEE 480
OY      421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERKFEVPEWTGSEE 480
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      661 PITNAGNFIKNTIGVGFSTGSRVYLASVNGDQRTGPNQFPLVYFVGLYGOQTRGT 720
OY      661 PITNAGNFIKNTIGVGFSTGSRVYLASVNGDQRTGPNQFPLVYFVGLYGOQTRGT 720
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780
OY      781 GTPESRSIFQSGYSDNTYEXYQSLGFDGIRNNLVNGVYASSFLNSNRPNGLEMLIA 840
OY      781 GTPESRSIFQSGYSDNTYEXYQSLGFDGIRNNLVNGVYASSFLNSNRPNGLEMLIA 840
OY      781 GTPESRSIFQSGYSDNTYEXYQSLGFDGIRNNLVNGVYASSFLNSNRPNGLEMLIA 840
OY      841 TTYLSQIGLARTSCLPNOOPFGTTHOYISVSPDOFSIKNRTIFPGNOLWYFLETFNE 900
OY      841 TTYLSQIGLARTSCLPNOOPFGTTHOYISVSPDOFSIKNRTIFPGNOLWYFLETFNE 900
OY      841 TTYLSQIGLARTSCLPNOOPFGTTHOYISVSPDOFSIKNRTIFPGNOLWYFLETFNE 900
OY      901 NKSASYTLRLADSSNPDASSSEFSTSLIDVNEIGVILPLDNSFYTVNAAGNALFESSN 960
OY      901 NKSASYTLRLADSSNPDASSSEFSTSLIDVNEIGVILPLDNSFYTVNAAGNALFESSN 960
OY      901 NKSASYTLRLADSSNPDASSSEFSTSLIDVNEIGVILPLDNSFYTVNAAGNALFESSN 960
OY      961 PGSPGTYAVNTFNONLSDIAFEGSAGKYSDFMGTIOFKPDEYLIQNGFTSOVARNFT 1020
OY      961 PGSPGTYAVNTFNONLSDIAFEGSAGKYSDFMGTIOFKPDEYLIQNGFTSOVARNFT 1020
OY      961 PGSPGTYAVNTFNONLSDIAFEGSAGKYSDFMGTIOFKPDEYLIQNGFTSOVARNFT 1020
OY      961 PGSPGTYAVNTFNONLSDIAFEGSAGKYSDFMGTIOFKPDEYLIQNGFTSOVARNFT 1020
OY      1021 NOSFLNSLVDEFTPANAGTYRVVDPDGNLNONPLKVOIQYLDGKTYDALKNNNLVT 1080
OY      1021 NOSFLNSLVDEFTPANAGTYRVVDPDGNLNONPLKVOIQYLDGKTYDALKNNNLVT 1080
OY      1021 NOSFLNSLVDEFTPANAGTYRVVDPDGNLNONPLKVOIQYLDGKTYDALKNNNLVT 1080
OY      1081 FSYNNFALPSWVPPAIGSTLGIILMIIIGLAIIGPLRAQRKLODDGFKTFKRVDTL 1140
OY      1081 FSYNNFALPSWVPPAIGSTLGIILMIIIGLAIIGPLRAQRKLODDGFKTFKRVDTL 1140
OY      1081 FSYNNFALPSWVPPAIGSTLGIILMIIIGLAIIGPLRAQRKLODDGFKTFKRVDTL 1140
OY      1141 TAAVGSYKRIITOTJANVKKRPAALGAGSGDKKPAAPAKPAAPKSPAKSPAKPTG 1200
OY      1141 TAAVGSYKRIITOTJANVKKRPAALGAGSGDKKPAAPAKPAAPKSPAKSPAKPTG 1200
OY      1141 TAAVGSYKRIITOTJANVKKRPAALGAGSGDKKPAAPAKPAAPKSPAKSPAKPTG 1200
OY      1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224
OY      1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224
OY      1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224

```

RESULT 2
US-10-131-591A-25

```

Sequence 25, Application US/10131591A
Publication No. US20030059799A1
GENERAL INFORMATION:
TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
FILE REFERENCE: J209
CURRENT APPLICATION NUMBER: US/10/131,591A
CURRENT FILING DATE: 2002-08-15
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 1224
TYPE: PRT
ORGANISM: Mycoplasma gallisepticum
FEATURE:
OTHER INFORMATION: MGC3 encoded by mgc3 gene
US-10-131-591A-25

Query Match      100.0%; Score 6413; DB 15; Length 1224;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1224; Conservative 0; Mismatches 0; Gaps 0;

OY      1 NMSKRLKSYTLIGLAVFALGASFGKSDKSDNTQLVNQARTLDANSVRLAGLG 60
OY      1 NMSKRLKSYTLIGLAVFALGASFGKSDKSDNTQLVNQARTLDANSVRLAGLG 60
OY      1 NMSKRLKSYTLIGLAVFALGASFGKSDKSDNTQLVNQARTLDANSVRLAGLG 60
OY      61 NGSLEFNTVLRDVEDNFITAANGTIKIDSEFTKPLXGLDLSDDCGYKVKQIVSDYTSRN 120
OY      61 NGSLEFNTVLRDVEDNFITAANGTIKIDSEFTKPLXGLDLSDDCGYKVKQIVSDYTSRN 120
OY      61 NGSLEFNTVLRDVEDNFITAANGTIKIDSEFTKPLXGLDLSDDCGYKVKQIVSDYTSRN 120
OY      121 RFQROTRAVYALVNDVDEANVHLKRLINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGT 180
OY      121 RFQROTRAVYALVNDVDEANVHLKRLINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGT 180
OY      121 RFQROTRAVYALVNDVDEANVHLKRLINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGT 180
OY      181 TKFNFTNOTGGEIYVNDFTLDPAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSG 240
OY      181 TKFNFTNOTGGEIYVNDFTLDPAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSG 240
OY      181 TKFNFTNOTGGEIYVNDFTLDPAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSG 240
OY      241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
OY      241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
OY      241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFNFNTOTGGEIYVNDPI 360
OY      361 LDAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSGTNADGMDGCGNGOITNT 420
OY      361 LDAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSGTNADGMDGCGNGOITNT 420
OY      361 LDAPILPKDLHPDWNYLQIKILPNDVNTAVVWPVGRVSGTNADGMDGCGNGOITNT 420
OY      421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERKFEVPEWTGSEE 480
OY      421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERKFEVPEWTGSEE 480
OY      421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERKFEVPEWTGSEE 480
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVNSLSFGIDSIYIF 540
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      541 GTSLEPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTGTGTTTADTSSGSGAGT 600
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      601 GNTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONEI 660
OY      661 PITNAGNFIKNTIGVGFSTGSRVYLASVNGDQRTGPNQFPLVYFVGLYGOQTRGT 720
OY      661 PITNAGNFIKNTIGVGFSTGSRVYLASVNGDQRTGPNQFPLVYFVGLYGOQTRGT 720
OY      661 PITNAGNFIKNTIGVGFSTGSRVYLASVNGDQRTGPNQFPLVYFVGLYGOQTRGT 720
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780
OY      721 FWYGTYKLLNSPDIYLDSPRVGTETNOFRTSLITYPVMGYLTGEGARSSNTPIYRAQ 780

```

Qy	781	GDPESSISIOGSDNTYEXIOGVLEPDDIRNNLNVGVASSFLNSNRPNPNGLEMTIA	840
Dd	781	GDPESSISIOGSDNTYEXIOGVLEPDDIRNNLNVGVASSFLNSNRPNPNGLEMTIAA	840
Qy	841	TTYLRISOIGLARTSGLEPNQOPFGTTHQVIVSVSPGDQSSIKNRTITIPGNOIWFLEPTNE	900
Dd	841	TTYLRISOIGLARTSGLEPNQOPFGTTHQVIVSVSPGDQSSIKNRTITIPGNOIWFLEPTNE	900
Qy	901	NNKSSVYTLRLADSSNPDASSSEFPSTLIQVNEIGVILLPLDINSFTYVNAAGNALFSSN	960
Dd	901	NNKSSVYTLRLADSSNPDASSSEFPSTLIQVNEIGVILLPLDINSFTYVNAAGNALFSSN	960
Qy	961	PGSEGSTAVVTFEONLSDIAFEGSGAKYTSDFMGTOIQFPRDEYLIONGFTSOVARNFVT	1020
Dd	961	PGSEGSTAVVTFEONLSDIAFEGSGAKYTSDFMGTOIQFPRDEYLIONGFTSOVARNFVT	1020
Qy	1021	NOFLNSLIVDTPPANAGNTRVNVVDPGONLTNOMLPLKVOIQOYIDGKYYTDAKLKNNNLVT	1080
Dd	1021	NOFLNSLIVDTPPANAGNTRVNVVDPGONLTNOMLPLKVOIQOYIDGKYYTDAKLKNNNLVT	1080
Qy	1081	FSYNNFGALBPWVYPTAIGSTLIGLITAMIIIGLIGIPLRQORLQDKGFTTFFKKVDTL	1140
Dd	1081	FSYNNFGALBPWVYPTAIGSTLIGLITAMIIIGLIGIPLRQORLQDKGFTTFFKKVDTL	1140
Qy	1141	TAAVGSYKKIITOTAVNKKRPAALGAGSGDKRPAANAARPAAPAKSPARASSPAKPTG	1200
Dd	1141	TAAVGSYKKIITOTAVNKKRPAALGAGSGDKRPAANAARPAAPAKSPARASSPAKPTG	1200
Qy	1201	PKSGAPTKPTAPKPAAPKPTAPKE 1224	
Dd	1201	PKSGAPTKPTAPKPAAPKPTAPKE 1224	

```

RESULT 3
US-10-131-591A-78
; Sequence 78, Application US/10131591A
; Publication No. US20030055799A1
; GENERAL INFORMATION:
; APPLICANT: Nippon Zeon Co., Ltd.,
; TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
; FILE REFERENCE: J209
; CURRENT APPLICATION NUMBER: US/10/131,591A
; CURRENT FILING DATE: 2002-08-15
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 78
; LENGTH: 1042
; TYPE: PRT
; ORGANISM: Mycoplasma gallisepticum
; FEATURE:
; OTHER INFORMATION: Modified MGC3 antigen (M11-BTR)
US-10-131-591A-78

```

Query Match	81.1%;	Score 5201;	DB 15;	Length 1042;
Best Local Similarity	84.7%;	Pred. No. 0;		
Matches 1018;	Conservative	1;	Mismatches 21;	Indels 162;
				Gaps 1;

QY	23	GSASGFEQOSKSNNDOTLVNOARTLDANSVRLGIGONGSLFNTVLDDVDNEITTAAG	82
Db	3	GSASGFEQOSKSNNDOTLVNOARTLDANSVRLGIGONGSLFNTVLDDVDNEITTAAG	62
QY	83	TIIKLDSFTKPLYGILDSDDCGYKVKOIVSDYTTSRNRFDOROTRAYALLVNDENVH	142
Db	63	TIIKLDSFTKPLYGILDSDDCGYKVKOIVSDYTTSRNRFDOROTRAYALLVNDENVH	122
QY	143	LKRINTNSRIGNRNNNSKEVIGGVNDPAHWIRFTDGTKEFNTNOTGGEIVNDEILDAP	202
Db	123	LK-----	124
QY	203	ILPKDLHPDWNTNLVYQRIILPNDVNTAVVPMVYGRVSGTNADDGHEFCGNGOITWTDPIA	262
Db	125	-----	124

QY	263	QTKTTTNDQNSTSTNSGAMPEANRRYSOLNVRKIRKTSFOLDRIKNTNSRIGNRNNS	322
Db	125	-----RINTNSRIRIGNRNNS	140
QY	323	KEFVIGVDNPNHVRFTDDGFKFNFNTOTOGELVNDPILDPILPKDLHPMWILYIORK	382
Db	141	KEFVIGVDNPNHVRFTDDGFKFNFNTOTOGELVNDPILDPILPKDLHPMWILYIORK	200
QY	383	ILPMDVNTAVVPWPEVGRVSGTNADGMEFCGNGOITMTDPLAQTKTTTNDONPSTFNSGA	442
Db	201	ILPMDVNTAVVPWPEVGRVSGTNADGMEFCGNGOITMTDPLAQTKTTTNDONPSTFNSGA	260
QY	443	MPGANNRRDSOLANKHRIKTSFOLDEKFPVPEMFGSEENKIRITLALGSLPSNERRYILD	502
Db	261	MPGANNRRDSOLANKHRIKTSFOLDEKFPVPEMFGSEENKIRITLALGSLPSNERRYILD	320
QY	503	IPGTPPYVTLKEDSVNFSRLYLNSVNSLSFIDGSIYIFGTSSELPMLWSEPTRLSDLT	562
Db	321	IPGTPPYVTLKEDSVNFSRLYLNSVNSLSFIDGSIYIFGTSSELPMLWSEPTRLSDLT	380
QY	563	LNQVKTDDIEASSVDNGTNGTNTYADTSSGSAGAGNTNTNSQVYNSPTLMTYRSFG	622
Db	381	LNQVKTDDIEASSVDNGTNGTNTYADTSSGSAGAGNTNTNSQVYNSPTLMTYRSFG	440
QY	623	IDSKPPTSANKIDENMMADPNVIEARIVAEYRIAGLONEPIPTNAGNFIRNTIGGVFTSTG	682
Db	441	IDSKPPTSANKIDENMMADPNVIEARIVAEYRIAGLONEPIPTNAGNFIRNTIGGVFTSTG	500
QY	683	SRVVLASRYNGDQRPNTGFPFLVFGYLGYYOQRTGTGFWGYTKLNSNPDYDLDBRV	742
Db	501	SRVVLASRYNGDQRPNTGFPFLVFGYLGYYOQRTGTGFWGYTKLNSNPDYDLDBRV	560
QY	743	GTEINOPFRSTLTYPVMGGYLTGEGARSEFSTPIYRAGDTPPEBSRSTFQSGSYDNTYXI	802
Db	561	GTEINOPFRSTLTYPVMGGYLTGEGARSEFSTPIYRAGDTPPEBSRSTFQSGSYDNTYXI	620
QY	803	QSVLGFDDIRNNLNVWVKASSFLLNSNRPNPGLMIATYLRQIGLARTSGLPNQPF	862
Db	621	QSVLGFDDIRNNLNVWVKASSFLLNSNRPNPGLMIATYLRQIGLARTSGLPNQPF	680
QY	863	GTHHOVISVSFGDDPFSSIKIRITRTPPGNQMLYFLFTNENKSSYYTLRLADSSNDPDASS	922
Db	681	GTHHOVISVSFGDDPFSSIKIRITRTPPGNQMLYFLFTNENKSSYYTLRLADSSNDPDASS	740
QY	923	FSPSLIDVNEIGVILLPLDNSFPTVNAAGVVALFSSNPGSGSTAVNFENONLSDIAF	982
Db	741	FSPSLIDVNEIGVILLPLDNSFPTVNAAGVVALFSSNPGSGSTAVNFENONLSDIAF	800
QY	983	EGSAGAKYSDPEWGTIOEKFPRDEYLLONGFTSOVARFNVYNOQSFNLSDYFEPANAGTNRV	1042
Db	801	EGSAGAKYSDPEWGTIOEKFPRDEYLLONGFTSOVARFNVYNOQSFNLSDYFEPANAGTNRV	860
QY	1043	VVDPDGNLTJNOLNPLKVOIYQLDGKYYDAKTLKNNLVTFSYNNFNGALPSSVVPYALIGSTL	1102
Db	861	VVDPDGNLTJNOLNPLKVOIYQLDGKYYDAKTLKNNLVTFSYNNFNGALPSSVVPYALIGSTL	920
QY	1103	GILAMITLIGLAIGILPRAOKRLDQKTKTFKRVDTLTAAVGSYKKIITOTANVKKRP	1162
Db	921	GILAMITLIGLAIGILPRAOKRLDQKTKTFKRVDTLTAAVGSYKKIITOTANVKKRP	980
QY	1163	AALGAGSGDKRPAALAAKPAAPAKFSAPKASSPAKPTGKSGAPTPTAPKPAAPKPTAP	1222
Db	981	AALGAGSGDKRPAALAAKPAAPAKFSAPKASSPAKPTAPKSGAPTPTAPKPAAPKPTAP	1042
QY	1223	KE 1224	
Db	1041	KE 1042	

RESULT 4
US-09-882-227-522
; Sequence 522, Application US/09882227

Publication No. US20030158396A1
 GENERAL INFORMATION:
 APPLICANT: Kleantous, Harold
 APPLICANT: Al-Garawi, Amal
 APPLICANT: Miller, Charles
 APPLICANT: Tomb, Jean-Francois
 APPLICANT: Oomen, Raymond P.
 TITLE OF INVENTION: Identification of Polynucleotides
 TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
 TITLE OF INVENTION: Genome
 FILE REFERENCE: 06132/04/002
 CURRENT APPLICATION NUMBER: US/09/882,227
 CURRENT FILING DATE: 2001-06-15
 PRIOR APPLICATION NUMBER: US 08/902,615
 PRIOR FILING DATE: 1997-07-29
 NUMBER OF SEQ ID NOS: 638
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 522
 LENGTH: 2893
 TYPE: PRT
 ORGANISM: Helicobacter pylori
 US-09-882-227-522

Query Match 3.5% Score 223; DB 12; Length 2893;
 Best Local Similarity 20.5% Pred. No. 7.8e-07;
 Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;

14 GGLAVFG-----ALGSASFGKQSDKSNQNTQVNOARTLDANSYRLAGLQNGSL 64
 287 GGTYTFNKEFSATNNTFFSGSFNFKVSSFN-GTFSNASTYFD-NQATFQNSPFGGT 344
 65 F-----NTVLRYDNDNEITANGTITIKLDSFTKPLGLDSDCCGKYKQVSD- 114
 345 FTENNQPTNNAHQIQONSSFQGNATITK-----GVNFQQAFFNS 387
 115 ---YTSRNRDQOTRAYYALLVNDANAHAKRINNSNRIGNNNNSKFIYGV---- 167
 388 NHQILIONASFNNAFTNNTGKITTEKQSEN---NTFTSYDITNMS--VAGVYLSG 441
 168 -----DNPAHVIRFDQTKFNTN-----OTGELYNDIILAPL--LP 205
 442 KNDLNGSTLDGSSKITLA-QGTFNLTSLGSEKSYTILNNSGGIYISMLNHNGLT 500
 206 KDLHPWYNLYIQRKILNDVNTAVVPPVGRVSGTANADGCFDGCQITNTDPIAQT 265
 501 SALKTN-----ESLSNPOSFOGLWDITTYGV-----TGQLLN-ENATFSK 541
 266 TTIDNQPSTFNSGAMPANRRYDQLNVKHRI-KTSFOLDERINTNSNR-----GNRN 319
 542 PIDSSPSKSTNSTQV-----YGVGYKIGDTIKLOETESHNSIITIALESSTYT 591
 320 ---NNSKRVYIG-----VDNPAHVIR--FTDDGKREFTNQTQGEIYND-- 358
 592 PPVINGSKFDLSASYINADMPWYDHKTYIKSONTFESGTYIPLSQVQINGSTNSPK 651
 359 -FLIDPPIPKDLHPM--YNLYIQRKILNDVNTAVV---PPVGRVSGTANADGCF 410
 652 TFSANGSNLYIGINSWTDHNVSSGTVSFGDTSGALNGHGPMPYVQCGTGT--NCTY 709
 411 DC-----GNCQITN-----TDPIAQTKTTDQONS-----TENSAMP 444
 710 SAYHVITIANLRSGNRIGTGGAANLIFNGVDSINIANATITQHNAGIYSSMTFTSQSD 769
 445 GANN-----RYDSOLNVKHKITSFOLDEREVEPMTGSENNKTRILATGSLPSNERVI 500
 770 NSONLNGLSNGKLSVGTFTFNPAKDGKIFNAGQAVENTNFN----- 814
 501 IDIFGPQVTLKEDSVAVFSRLYIN-----SYNLSLIGDSIYIFGSELPLM 549
 815 ---GGSYQ--FSGSLNFSNNQFNNSGFEISAKNNSFNANNSASFENNSATTSF 869
 550 YSEPTRLSDITALNOKTIDIEASFDNGTGTNGTTTADTSSGSGTGAGTGTNT----- 603

870 VGDEFNANSNL-----QI-----AGNAVFNSTNGSQNTANFNNGSVNISCNATFDNVY 919
 604 ---TNTS-----QTVSNPT---INTYRFG-----IDSKP-----TSA 630
 920 ENGPTNISVKQVVLNNTTKLNLAAPLFGCGITFENANSHVINIAESTINGNTILVSS 979
 631 NKIDETNADPNVTEARLYAERIGIONEIPITNAGFINNTGCVGTSGSVYLRAS 690
 980 KEIEYNNAFSNNLMQILNYQGH--GASEKLVSAGNGVDV-----YS 1022
 691 YGDDRPNGRQPFV-----YVEGYLG-----YQRTTG--TWYGYTK- 727
 1023 FNNQ---TYNROVESQNSISIRLGVNMYVDYDMKSDHLTYQNALGEMTPNPN 1079
 728 ---LNNSPY-----DYLDSPRVOTENORRRTSLYPVAGGYLTEGARSPSNPEYR 778
 1080 NLGNANNTLYYYDKSIDFYASGTLFTAEFSQI-----FTGQNSAIVFGAKSIWT 1130
 779 AGDPRPSRSIFQSGYSDNT-----YE-----YIQ----- 803
 1131 SLSDAPQSTTILRFQ--DNKAGASNDASGHCWNLQCIQITGHYEAOKIYTGSIENGR 1188
 804 ---SVLGFQIRN-----NLNVGYKAS--FLNSRPNPGLNEMIAATYLR 846
 1189 ISSGASINENGLQILLTNATLYNRAAGTSSMNFIS-----NGANIOANSYFID 1242
 847 QIGLARTSGILPNOQPGTTHQVYISPPGQFSSIKNRIPIFGNOLWFLFTENKSS- 905
 1243 D--TAQNGNPN--FSFNALNIDFSSNSFRGVGQTSVEFKFNKNAISTFNTSNLSSG 1297
 906 YVTLR-----LADSSNDPSSFSPTSLIDVNEIGV-----LPILDN 943
 1298 LYQMAQASVLEDSNLSVSVG---TSSIKANALINLSQNASINASHSTLEGLDGNVDT 1354
 944 SFY-----TYNAGNVAL--FSSNPGSPGTYT---AVN-----TFNQLSDIAFEGS- 985
 1335 SSLMNGSTINVSNNATINDYASLIASNGSHLNFNGAVNNSANTTSNLSYFPGAV 1414
 986 --GAKY-----TDFEQTQIFKPDDEVLIQNGFTSQVARNFYVNOSEFLNSLVDTPPANG 1037
 1415 SLGGQFNLSSNLSLDPQGS-----SALTSNARFQYDN-AFSQSPITFHOA--- 1459
 1038 TNYRVVYDDGNLTNOMLPLKVOIYLDGKRYDAKLNKNNLYFYSNNGAL 1089
 1460 LDKAPLPLGNNLNP-----NSSVLD-----LKNSQLV--FEDQSL 1496

RESULT 5
 US-09-813-214A-9
 Sequence 9, Application US/09813214A
 Patent No. US20020177200A1

GENERAL INFORMATION:
 APPLICANT: Tucker, Kenneth
 APPLICANT: Piosella, Laura
 TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE.
 TITLE OF INVENTION: SEQUENCE AND USES THEREOF
 FILE REFERENCE: 7969-089-999
 CURRENT APPLICATION NUMBER: US/09/813,214A
 CURRENT FILING DATE: 2000-03-20
 PRIOR APPLICATION NUMBER: 08/968,685
 PRIOR FILING DATE: 1997-11-12
 NUMBER OF SEQ ID NOS: 20
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 9
 LENGTH: 2122
 TYPE: PRT
 ORGANISM: Moraxella catarrhalis
 US-09-813-214A-9

Query Match 3.2% Score 205.5; DB 10; Length 2122;
 Best Local Similarity 19.5% Pred. No. 9.6e-06;
 Matches 282; Conservative 156; Mismatches 502; Indels 507; Gaps 70;

QY	24	SASGEFQOSKSNNTQLVNQARFL-----DANSVRLAGLSON-----GSLNPTVLAD	71
Db	482	AATGG--QLKRVNQTASALQTFYVKKVADKNGDANDASKITTVKKNKPCDTQYNT----	535
QY	72	VDDNFITPAANGTIIKLDSPFKPLGLDLSDDCGGYKQOIVSDYTTSSRNPEDQROTARAY	131
Db	536	-----LKL-----KGEGVDYTFTETG-----IYTF	556
QY	132	ALLVDEDAVNHAKIINTNSNRIGRRNNSKEVIG-----GYDNPAYIR	175
Db	557	GLANONNOLTVGNSTLINDGLSVKNTNSNQIOVADGITTETDISNSKRGAIEN---TTR	613
QY	176	FTDDGTFKNFTNTOGGEIVNDFILDAFLLPRKHDPWYNLYIQSKLILPNQVNTAVPWPV	235
Db	614	ITRDG--IGFPAANTGS-----LDAN-----KPLPITPGIN-----A	642
QY	236	GRVSGTNADDMGPFDCNG-----QITWDEIAOTKTTD-----NQNSTFNSGAMP	282
Db	643	GGRKELTVQSAINPATNGGOLDPFNNRLSTATETKSGSATIKDLNLSQVLEITPAAGDGP	702
QY	283	GANNRY-----DSOLNVKRIKTSFOLDERINTNSNR	314
Db	703	NWTKKLEILLKVGKTTADDLTKNNIGNVADSTDSNLSYK--LAKTSLDLD--AVNTKTLT	760
QY	315	IGNR-----NNSKEVIG-----GYDNPAYIRFD-----DG	342
Db	761	ASDKVAVYDGSNNRTAKLNGDLTFESKONTGATPATNSKITGVDG---LKFIDNGIALDG	816
QY	343	TKFNFNTQOGEIYNDFILD--ABILPRDLHPWYNLYI--ORKLIPNDVNT---AVPWP	396
Db	817	TTY--ITMDKYGFAODSDLOKSKRYLKD-----KLKGVELTTNGINAGGAITGLS	869
QY	397	VGRVSGTNADG-----MFDCNGQITWTD--PIAOTKTTDNO	433
Db	870	NTLTDATNATGHTYQLGIVDSTDKTRAASIGDVLNAGFNLNKNGDAKDFVSTDEYD--	927
QY	434	NPTFNSGAMPGANNRRDSOLNVKRIKTSQOLDEKEYVPWMTSSEKKNITRLATGSLP	493
Db	928	---FINGNATTAAYTYDGKAS-----KVAADVVDGTHLHTLADAGKNNDIGKTYTLT	978
QY	494	SNERWYILDIPGTFQVYTLK---EDSVNVFSLRLNSVNSLSFGDSIYFGTSELPLW	549
Db	979	KT-----DAKGDKAINFSVNSGDDKALINKDADNINLTIA--GEIRNTGTAD-----	1025
QY	550	YSEPTRLSDLTALNOVKTDDIEASSTDNCTTTNGTTTADTSAGSGAGTGNNTNSOT	609
Db	1026	-----TALQTFQVKKYK---ENGDDN---DADTIT-----YGDAKTNO-	1059
QY	610	VSNTPLTWYRSFGIDSKRPTSAKIDEITNMADPNVIEARVIAEYRIGLONEIPTITNAGFI	669
Db	1060	-----VMTLKKKGKNGLDIOQTK-----DGTVTGELINQSGIKAGANNNTLNNNGLSI	1106
QY	670	RNTIG-----GVGFTSTGSRAVLASVNGDORPFGNPOFPLYVGYGYGOOTRGT	720
Db	1107	KNTAGNNOIQVGADGVFAKAYKNNGV--GAGIDGTYTTRTRDIEGACTNGSLDKSKPHLSK	1165
QY	721	--FWYGYTKLLNNSPYDVLDSPRVGTETNOERRTSLIYPVNGYL---TE-----EGA	768
Db	1166	DGIAGAKKKTIN-----IOSGEIAONSND-----AVTGKRIYDLKTELEMKISSTA	1211
QY	769	RSEFNTRYIARAQGTPEPSRSIFQCGSYDNTRYEQSVLGF---DGIRNNLN-----VGK	820
Db	1212	KTAONSLSHEFSVADQGNFTVSNPYSYDTSKTSDYTFTEGEGNITTKRVKGVARRVIGD	1271
QY	821	ASSEFNLSNR---PMPNGLEMIATUTYLSQIGLARTSGLPNQOPFGTHOIVISVPGQF	877
Db	1272	OTKRLTTPKPLTVGNNGKGIY-----IDSQNGQWTTIGLSN-----	1307
QY	878	SSIKNIERTIFGNOLMYFLFTNENNNKSVYTLR---LADSSNDASSFSFPTSLIDVNE	933
Db	1308	-TLANV-----TNDKGSVTRTEGQIKIKDEDKTRAA-----SIVDLS	1344
QY	934	IGVIL-----PLLD--NSFTYVNAAGNVALFESSNPGSPQSYTAVNTFNONLSDIAPESGGA	987

```

Db      1345 AGFNLQNGEAVDVSSTFYDDYVFPADGNATTAATKATYTDTSKTSKYYVYDVNDDTTIEYKDK 1404
QY      988 K-----YTSDFNGTIOFKRDEYLIQNGFT-----SOYARNFVTNQSFLNSLVDFTP 1033
Db      1405 KLGKVTTLTLSTGCA-----NKFALNSQATGDALVAKASDVAHLNLTLSGDIQTAKGASQ 1459
QY      1034 ANAGTNRVYVDPDGNLTNOMLPKVOIOYLDGKYYDAK-----LKNNNLVYFYSY 1083
Db      1460 ANSSAGY--VDADGN-----KVITYSDTNKYYAKKNDGTVDKTEVAKDKLVAAQ 1508
QY      1084 NFGCALSWVVPYPAIGS-----TLGILAI 1107
Db      1509 TPDCITLQOMNVKSYTNKEQVNDANKKCGINEDNAFVYGLEAKASDNKTKNAATVYGDINA 1568
QY      1108 MIIIGLAIGIPRAQRKLQ--DKGFYTFEKV-DTLTAAGVSYKKIITQ----- 1154
Db      1569 V-----AQPLTFEAGDYG--TTAKKLEBELFLIKGGQGDYTKMLTDNNIGVAGT 1614
QY      1155 ----TANYKKKFPALGAGKSG----DKKPAALAKPAADPAKSPAPASSPAKFTGKR-----S 1203
Db      1615 DGFVYKLAKLDTLNLNSYNAGTGKIDDKGVSPVDSGQAKANTPVLASANGLDLGKRVISNV 1674
QY      1204 GAPTPT 1210
Db      1675 GKGTGPT 1681

RESULT 6
US-09-815-242-12713
? Sequence 12713, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Kari L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes In
? FILE REFERENCE: ELITRA.011a
? CURRENT APPLICATION NUMBER: US/09/815.242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 12713
? LENGTH: 2344
? TYPE: PRT
? ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 3.2%; Score 205.5; DB 9; Length 2344;
Best Local Similarity 19.2%; Pred. No. 1.1e-05;
Matches 226; Conservative 167; Mismatches 483; Indels 301; Gaps 50;
2 NISKRLKSYTL-----IGGLAVFSGALSGASGFGQSDK-----SNDNTQLVN 43

```


Db 1564 TSGNATATSFNTVAKP-----LRDKYRWGTS-----STAANPYRIANISMDA-- 1605
QY 325 VLGVDNPAHV--IRFDGDKFNFNTNQGEIYVDFILDAPILPDLHPDMYLNLYIQRK 382
Db 1606 TVSQADQVTAIINSLMETETVPNKNMATAANETISKTV-----SNVSRGTNNANVTY 1658
QY 383 ILPNVNTAVVPMVGRV-----SGTNADCMFCOGN----- 414
Db 1659 VTYQDGTSTVTPVAKHVLPELVASHYTVQGDPAFGSSHAD-YFKLSNGSAIPDXT 1717
QY 415 -----GQITNTP-----PIAOKT-----TTDNONPSTFNSGAM- 443
Db 1718 ITWVSGAENKDKRIGEDINVTANILIDGETPIPKTATYKVSVPVHFETNGAYF 1777
QY 444 PCANRYDSQLVNKHAKIKTSFQDEKVFYVPEWTGS-EKKNTIRLATGSLP---SNERTW 499
Db 1778 PGVSDVYDAKQYVK-PVNDSPWQOMAGRMFOFTNSGSPKSDVVGISTRIIRIYTYDHQIQ 1836
QY 500 IADI-----PGTPQ-----VTLKEDSVNFSRLYLSVNSLSPIGDIYIFGTSELSLM 549
Db 1837 IIKILAKVPRDPRIDGNSVTYKAGLTN--QOKIKNNVSSS-----STKLFKADNTP--- 1887
QY 550 VYSEPTRLSDLTALNQVKTDDIEASSTONGTTTNGTTTAD-----TSSGSGAGGNTT 604
Db 1888 -----LITTN-----TTGSGNTAVVTVSDALPNCVIRKASSITNNNTY 1926
QY 605 NTSQVYSNPTLNTYRFSFGIDSKPTSAKIDETNMADPNVTEARIYAE-----YLG----- 655
Db 1927 YTTDDEGRAIDVTRNESVDSNDSATVYTPPOLA--TTEGAVFKGDDGDFGVHEER 1983
QY 656 ION-----EIRTNAGNFIRNTIGVGFSTGSRVLRASVNGDOR-----P 697
Db 1984 IQNPRGATVAMHNDNT-----WKNTVNTIKTA-----VTLPSQSGRANVEPVKYP 2034
QY 698 TGNFO-PPLVYFGYLGQOTRTGTETWYGYTKLINSPYVLDSPRVGETN----- 747
Db 2035 VANAKAPSRDVG--QNLNGT-----DAINYTFD-----PNTWNGITAMANR 2078
QY 748 -----OFRRSIATP-----VMGQYLEGARSFSTPIRAGDIPESR 787
Db 2079 QOBNNOAGVQHLNDVTPGTTAKRVVYVNAVQFE-----FPQSTYTTVGGT--- 2129
QY 788 SJFOSGSDNTYEYIQSVLGF--DGIR--NNLNVGVKASSFNSRPNNGLEMIATY 842
Db 2130 --LANGTOASGYAHMQNANGLPDGFYKMNNAATCTNDANMAAMKPN---AAKYVNAK 2184
QY 843 YLSQIGLARTSGLP-----NOQ-----FGTHOVISYSPQDQFSSIKNTITPPGNO 891
Db 2185 YQVITYNGHFAATSLPAKFVVKDVQPAKPTVTETAGACATITTPG---ANQTVNT-HAGNV 2239
QY 892 LMV-----FLFTNNKNS-----SVYTLR-LADSSN--PDASSSEFPTSLI 929
Db 2240 TTYADKLVTIKRNGNVYTTTTRNNNSPWKKEASATVAGIAGTNNQITVYAGTFNADPT 2299
QY 930 DV-----NEIGVILPDLNSEFYVNAAGNVALFSNPGSPGSYTAVNTE 973
Db 2300 QVAVATGSETISDEQRSDDFYVAPQPHANAKIKINGHIDITPNP-----SGHLINP- 2354
QY 974 NQNLISIAFE--GSAKAT-----SDFWGTIQFPRDEYLIONGFTSOVAKRPVTN 1021
Db 2355 --QOAMDIAYTERKNGNAEHSKITINVVGONNOM-TIANKPD-YVTLDAHKGK--TFENAN 2409
QY 1022 QSFNLNVLDTTPANACTNY-----RVVYDPPGNLTNQMLPKYQI 1061
Db 2410 TYKPNASAITTP-KACTGHSASSNPSTLAPATHYVNTETIYKDYGSVNTAETINNAVQV 2468
QY 1062 QYLDGKYDAKLKN-----NNT-----VTFSYNN 1085
Db 2469 ----ANKRTATIKNGTAMPNTNLAGSTTTTLPVTVYND 2502

RESULT 8
US-10-238-075-749

Sequence 749, Application US/10238075
Publication No. US20030148324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are 1st
TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of 1
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
CURRENT FILING DATE: 2002-09-10
PRIOR APPLICATION NUMBER: 0003145
PRIOR FILING DATE: 2000-03-10
NUMBER OF SEQ ID NOS: 1576
SOFTWARE: PatentIn version 3.1
SEQ ID NO 749
LENGTH: 1778
TYPE: PRT
ORGANISM: Escherichia coli
US-10-238-075-749

Query Match 3.1%; Score 197; DB 12; Length 1778;
Best Local Similarity 19.7%; Pted. No. 3.1e-05;
Matches 259; Conservative 150; Mismatches 475; Indels 428; Gaps 69;

QY 2 NISKKLSXYTL-----IGLAVFGALGSASFQKQSDKSN----- 36
Db 294 SIASKVNSIALGNSLSGGENALALGEGSAGGSNSLAFGSGSRANGNSVAIGVAAAA 353
QY 37 -DNTQVLVNAFTLDA-NSVRLAGLGONGSLFNTLRVDNDNFITANGTIIKL--DSFTR 92
Db 354 TDSVAIIGAGSTFDASNTFVSGNSATRKRIIVNNAAGAIISNTSDAINGSQLTYISVAK 413
QY 93 PLYGDLSDDCGGYKKQIYSDVDTTSRNRFDDQOTRAYALLVNDANHLKRIINNSR 152
Db 414 RLGG-----GATVGSQDGV-----TAVSYAL-----R 435
QY 153 IGRNNNSKFVIGVDNPAHVIRFTDDGTF-----NFTNQ-----TOGEIYN 195
Db 436 SGTYYNVGD-ALSGIDN--NTLQMNKTAGAFSANGHANGNTKITTNAKGVASATSDVYN 492
QY 196 DFLDAPILPKDLHPMYNLVYQKILPNDVNTAVVPMVGRVSGTNADDMFCNGQI 255
Db 493 GSOL-----YDLQ--ODALLMN--GTA-----FSAHGTETASKITNTVATGML 531
QY 256 T--NTDPI--AOKTTTNDONPSTFN-----SGAMPGAN-----NRYDSOLNV 294
Db 532 TAGSTDAVNGSQLKTINDVNTTNTIATNTTITNLVTDVANGSGDSSLMKAKAFAFSA 591
QY 295 KRIKTSFOLDERINTNSRIGRNNNSKFVIGVDNPAHVIRFTDDGTFKFNFTQOGE 354
Db 592 AHGTEATSKI-----TNVT-AGMLTAGSTDAVNGSQ-----LKTNDNVNTTNTIATNT 640
QY 355 IVNDFILDAPILPKDLHPMYNLVYQKILPNDVNTAVVPM--PVGRVS---GTNADDMG 409
Db 641 TINTNLTLDA-----VNLSGDSLLMKNTAGAFSAHGTDATSKI 679
QY 410 FDCGNGOIT--NTDPI--AOKTTTNDONPSTFN-----SGAMPGANRYDSOLNVK 457
Db 680 TNYTAGNLTRAGSDAVNGSQLKTINDVNTTNTIATNTTITNLVTDVANGSGDSSL- 738
QY 458 HRIKTSFQDEKFEVPEWGTGSEENKNTIRLATGSLPSNEREYIILIPGPQVTLKEDSVN 517
Db 739 NKTAGAFSAAH-----GTDATSKITINVKADILTAGS-----TDVANGSQLKTINDVNS 786
QY 518 VFSRLYLSVNSLSIFGDSIYITIGTSELSPLMYSPPTRLSDLTALNOVKTDDIEASST 577
Db 787 TNYTNTNLTLDAVANGSGDSSLNWN--KTAGAFSAHGTDATSKITINVKAGDLTAGSD 842
QY 578 --NGT--TTNGTTTADTSSGAGTGTNTTSQTV-----SNP 613
Db 843 AVNGSQLKTINDVNT-----NTNITNLVTSVGDYKDKDSSLMKAKAFAFSAHG 892
QY 614 TLNTYRFSFGIDSKPTSAKIDETNMADPNVTEARIYAEYRLGIQNEIPTNAGNFIKNTI 673

[illegible]

Db 1044 GN-----DGGSPFN-----INMGKMLEDR---SDMIYNTMNVG- RNMMLAL 1083
QY 1095 -----PRAISSTIGILAIMIILGLAT-----GIPLRA-----ORKIO 1126
Db 1084 RPSRKFTNKHFFYPALIGVCMILSALVSFMPAVNTKHNIKLSATNEDLHKEIYNRKILA 1143
QY 1127 DKGFTEFKKVD 1138
Db 1144 EKALAESQERLE 1155

RESULT 10
US-10-172-502-4
; Sequence 4, Application US/10172502
; Publication No. US20030185833A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES.
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2283
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-4

Query Match 3.0%; Score 192.5; DB 12; Length 2283;
Best Local Similarity 18.7%; Pred. No. 9.9e-05;
Matches 219; Conservative 167; Mismatches 494; Indels 293; Gaps 47;

QY 2 NISKKKKSYTL-----IGGLAVGALGASFGFGQSDK-----SNDWTOLV 42
Db 70 SISKKMTGGLKTTAVIGGAFIVNMLHDQ-AFAASDAPLTSELNQTSEVGNQNSTTIE 128
QY 43 NQARTIDANSYRLAGLQGNKSLFNLYLRVDYDNFTIANGTITIKLDSFTKPLYLGLDSD 102
Db 129 ASTSTADSTSVTNSSSVQTSNSDITVSEKSEKVTSTNSTNSQOEKLT-----178
QY 103 CGGYKQAYSDYTTSRNRPDQROTAYVALLVNDANVALLKRIINTNSNIGNRNNKSF 162
Db 179 -----TSESTSSKNTTSSSDTKSVASTSEBP-----INSTNOSTSNNTSOS 223
QY 163 VIGGVNPAHVIRETDDGTRKFNFTQGEIVNDPILDKLHPDWYNYIORKIL 222
Db 224 T-----TPSSVNLNKTSTSTST-----TAVKLR-----FSLAMSTFA 258
QY 223 PNDVNTAVVPMVPGVGRVSGTNADGMDCCGQITNDPIAQTITTTDNONPSTF--NSGA 280
Db 259 SAATTTAVT-----ANTTITVKNKDNLKQYMTTSGN--ATYDOSTGI 296
QY 281 MGCANNRYSQ-----LNVKHKIKTSFOLDERINTNSRIGNRNNNSKFIYGVNDNPAHVI 336
Db 297 VITLQAYISQKGAITLGTRLDSNKSFTHSKVLNGLNKYEGCHN-----GG-----341
QY 337 RPTDDGTRKFNFTQGE-----IVNDPILDAFILPKDLHPDWYNYIORKIL 384
Db 342 -----DGIGFAFSPGVGELGNGAAYIGISLNAF-----GFLDLYH-----NTSK 384
QY 385 PNDVNTAVVPMVPGVGRVSGTNADGMDCCG--GOITNTDP--IAQTKTTDN--QNPSTFN 439
Db 385 PNSAKAKA-----NADPSNVAAGGAFGAFVYTTDSYGAATYTTSSSTADNNAKLN 432
QY 440 SCAMPGANNRYSQGLANVHKRIKTSFOLDEKFFVPEWTGSEENKNI-----484
Db 433 --VQPTNNTFODPDIN-----INGDTRKVMTKTAGOTWTRNISDWIAKSGTTFNLS 482
QY 485 -----TRLA-----TGLSPSNERYWILDI--PGTPQVTLKEDSVNVFSRLYL--N 535

Db 483 MTASTGATNLDQVQGFTEYTESAVTQYR--VDVTTGCKDILPKRTYSGNDQVYITDN 540
QY 526 SVNSLSFTG-----DSIV-----IFGTSLEPLSMYVSFPPLRLSDILANOV---566
Db 541 QQSALYAKGYNYSVDSYASTYNDTNKIVTKMNAQOSVYVYETPDVAKAPVYVGNQITIEV 600
QY 567 -KT-DDIEASSTNGTNGTNGTTTADTSSGTSAGGNTNTSQTSNPLNTYRSPGID 624
Db 601 GKTMPVILTTDNGTIVTNTVYTGSLSDSADNIIIGTPTKIGOSTIVYVSTDOAN 660
QY 625 SKPTSAKIDETMMADPNVI-----EARIYAERLQIONEIPIT-----NAGNFIRN 671
Db 661 NKSTTFTINVDTPATPTVPIGDQSEVVS-----PLSPIKIAIQDNAGNAVYN 710
QY 672 TI-----GGVGFSTGSRVYLKASVNDQRPCTGNFQPLVFGGLGQRTGTGFWGTYK 727
Db 711 TVYGLPSGLTFSDTNNTI-----SGTPNTIGTSTISIVSDASGNKTYTTF---KYE 759
QY 728 LUNNSPYDVLDSPRVGTEFNQPRFTSLT-----YPMGCVLTEGARFSMTPIYRAQ 780
Db 760 VTRNS---MSDSVSTSGSTQOSQSVSTSKADSQASTSTSGTIVSTASTSKSTSVLSL 816
QY 781 GDTPEBSRIFQSGYSDNTYEQYVGLGFDGIRNNLNVGVKASSFLNSRPNGLF---836
Db 817 DSVASAKSLSTESNSVSSSTSTSLVNSQSVSSMSDASAKSTSLSDSISNSSSTEKSES 876
QY 837 -----MTAATTVLSQGLARTSGLPNOQPEGTHQVYVSFGQFSSIKIRIPIFC 889
Db 877 LSTSTSDSLRTSTSLSDSLSMSTSGSLSKQSLSLSTSGSSYSTASLSDSTSN--AISTS 934
QY 890 NOLMYFLFTNE--NNKSSVTLRLADSNDPDASSSPSTSLIDVNEIGVILPLDMSFTY 947
Db 935 TSLSEASSTDSISISINSASASTSKSDSOSTISLSTDSKSKMSTSESLDSTSTIS 994
QY 948 VNAAGNVALFSSNPGSPGSYTAVNTFNQNLSDIAFEGSGAKYTSDFWGTIQKRPDEYLIQ 1007
Db 995 GSVSGSLINASQSVSTSTSDSMST--SEIVSD--SISGSLASDS--KSMVSASSMSTSQ 1051
QY 1008 NGFTSQ-VANFPTNOSFLNSLYDTFPANAGTN 1039
Db 1052 SGSTSELSDSOSTSDSKLSQSTSGSST 1084

RESULT 11

US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match 3.0%; Score 192.5; DB 12; Length 2468;
Best Local Similarity 20.2%; Pred. No. 0.00011;
Matches 308; Conservative 154; Mismatches 503; Indels 563; Gaps 84;

QY 61 NSGLENTVLKVDNDNFTTANGTIIKLDST---KPLYGLDLS-----DPCGGY 106
Db 559 DGTIVNVVARDAGNSSPPASVTVDAVAPATPTVDSNGTTLGTAEFGSSVTLTDGNGN 618

1583 VISGTEAGATVILFDGNPNPIGOVYADSGSNWSTTPGTPLANSVINALADPAAGNNS 1642
1583 VISGTEAGATVILFDGNPNPIGOVYADSGSNWSTTPGTPLANSVINALADPAAGNNS 1642
925 PSLIDVNEIGVLLPLDINSFYTVNA---AGNVALFSSNPGSPGSTYAVTFNQNLSDI 980
925 PSLIDVNEIGVLLPLDINSFYTVNA---AGNVALFSSNPGSPGSTYAVTFNQNLSDI 980
1643 PLS-ATVDSIAPAPVYIDPSNGSVINGTAEAGATVILFDGNPNP-----IGOV 1689
1643 PLS-ATVDSIAPAPVYIDPSNGSVINGTAEAGATVILFDGNPNP-----IGOV 1689
981 AFESGAGAKYTSDFWGTIQFPEDEYLIQNGFTSQVAFNEFTNQSFLNSLVD-----F 1031
981 AFESGAGAKYTSDFWGTIQFPEDEYLIQNGFTSQVAFNEFTNQSFLNSLVD-----F 1031
1690 TADSGN-----W---SFTPGPPLSNGTVYVNAVADPAAGNISGVPSTVDVAVATPVI 1740
1690 TADSGN-----W---SFTPGPPLSNGTVYVNAVADPAAGNISGVPSTVDVAVATPVI 1740
1032 TPAN---AGT---NRYVYV-DPDGNLTNOMPLKVOIOLDGKYDALKNNNLVTFXY 1083
1032 TPAN---AGT---NRYVYV-DPDGNLTNOMPLKVOIOLDGKYDALKNNNLVTFXY 1083
1741 DPSNGVELSGTAEPGVRYILFDGNGNIGQTL-----ADSS----- 1776
1741 DPSNGVELSGTAEPGVRYILFDGNGNIGQTL-----ADSS----- 1776
1084 NNEGALPSPWVPTAIGSTGLIAMIILGLAIGIPLRAROKLDDKG-----FKTTFKYV 1137
1084 NNEGALPSPWVPTAIGSTGLIAMIILGLAIGIPLRAROKLDDKG-----FKTTFKYV 1137
1777 GNMSTPG--FPLANGTVYVNAVA-----QDPAGNTSGPASTVDVTV 1815
1777 GNMSTPG--FPLANGTVYVNAVA-----QDPAGNTSGPASTVDVTV 1815
1138 DPLTAAGSVYKKITTOTANVKKRPAAL-GAGK-----SGDKRAAARAP----- 1184
1138 DPLTAAGSVYKKITTOTANVKKRPAAL-GAGK-----SGDKRAAARAP----- 1184
1816 APATVYVINSNGSVITGTAEVAKVILFDGNGNPIGETATDAGSNMTFTPTPLANGTV 1875
1816 APATVYVINSNGSVITGTAEVAKVILFDGNGNPIGETATDAGSNMTFTPTPLANGTV 1875
QY 1185 ---AKPSAPKASSPAKPTGPKGATPK 1209
QY 1185 ---AKPSAPKASSPAKPTGPKGATPK 1209
Db 1876 NAVAEDACNAGNAGSPASTT-VDSVAPAP 1902
Db 1876 NAVAEDACNAGNAGSPASTT-VDSVAPAP 1902

RESULT 12
US-10-193-764-63
; Sequence 63, Application US/10193764
; Publication No. US2003013943A1
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michael H.
TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
FILE REFERENCE: 1038-1239MIS
CURRENT APPLICATION NUMBER: US/10/193,764
CURRENT FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: 09/167,568
PRIOR FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 63
LENGTH: 1536
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-193-764-63

3.0%; Score 192; DB 12; Length 1536;
Query Match
Best local similarity 19.5%; Pred. No. 5.9e-05; Indels 458; Gaps 66;
Matches 266; Conservative 155; Mismatches 487;
27 FGEKQSD-----KSDNTOLVNOARTLDANSV-RLAG-LGQNSLFFNYLRVDNDFITA 79
27 FGEKQSD-----KSDNTOLVNOARTLDANSV-RLAG-LGQNSLFFNYLRVDNDFITA 79
105 FNIDNENVOFLQNNNSAVFNR---VTSNQISQLGILDSNGQVFL-----INFGITI 156
105 FNIDNENVOFLQNNNSAVFNR---VTSNQISQLGILDSNGQVFL-----INFGITI 156
80 ANGTLIKIDSPFKPLGYLIDSDCCGKYKVOIYSDTTSFN-REFDQRTA-----YIA 132
80 ANGTLIKIDSPFKPLGYLIDSDCCGKYKVOIYSDTTSFN-REFDQRTA-----YIA 132
157 GKDALIINNGFTAST--IDISNE-----NIK-----ANLFEEDTKDAIAEIVHGL 202
157 GKDALIINNGFTAST--IDISNE-----NIK-----ANLFEEDTKDAIAEIVHGL 202
133 LLYANDEANVHLKRITNSNRIGNRNNNS-----KFTIGGVNDPA-----HVIR 175
133 LLYANDEANVHLKRITNSNRIGNRNNNS-----KFTIGGVNDPA-----HVIR 175
203 IYVGKDGSVNLIGGKVKKEGVISVNGGSIISLAGOKITISDILINPTIYSIAAPENAVN 262
203 IYVGKDGSVNLIGGKVKKEGVISVNGGSIISLAGOKITISDILINPTIYSIAAPENAVN 262
176 FTTDDGKFEFTNOTGEIYVDEILDAPIILPKDHLHDWVNLVYQRIKILPNDVNTAVPWPV 235
176 FTTDDGKFEFTNOTGEIYVDEILDAPIILPKDHLHDWVNLVYQRIKILPNDVNTAVPWPV 235
263 LGDIFKGGNINVRATIIFNQCKLSDSVSKDKSG---NIYLSAKBEAEIIGVI----- 314
263 LGDIFKGGNINVRATIIFNQCKLSDSVSKDKSG---NIYLSAKBEAEIIGVI----- 314
QY 236 GRVSGTNADDCGMDGNGOITMTDPIAQTKTITD---NQNDSFENSGAMPANRRYSOL 292
QY 236 GRVSGTNADDCGMDGNGOITMTDPIAQTKTITD---NQNDSFENSGAMPANRRYSOL 292

```

Db 315 -----SAONQAKGKLMITGDKVTLKTGAVIDLSGKEGERTY-----LGDGERGGRK 363
QY 293 NKHKRIKTSFOLDERINTNSNRIGNNNNSKPIGVDPNPAHRIRTDGTCKNFNFNQ 352
Db 364 GIOLAKKISLEKSTINVSKEKGR---AIWMDI-----ALIDG--NINAQGS 408
QY 353 GEIVND-----FLIDAPILPKD---LHPDWNYLTIORKILPNDVNTAVVPWPV 398
Db 409 GDIAKTGFVETSGHDLFIKDNAIVDAKEMLDLPD-----NVSINMETAG 453
QY 399 RVSGTNADDMDCGCGQIT--NTDPIAQTKTTTTONONPSFTNSCAMPG--ANNR--YD 451
Db 454 R-SNISEDEYEGSGSASTPKRNEKTKTLTNTLE-----SLIKKTFVNITANORIYVA 508
QY 452 SOLNVKRIKTSFOLDERKFPYPEWT-----GSEENKNITRLATG-----SLPSMERW 499
Db 509 SGINLSNGSLTL-----WSEGRSGGVEINNDIT--TGDDTRCANLTIYSGGW 554
QY 500 I-----LDIPGTPVTLKED-----SVNV-----FSRLXNSVNS- 529
Db 555 VDVHKNISLGAOGNINIFAKODIAFEKGSNOVITGCGTISGNOGFRFNNSVLNGTSG 614
QY 530 LSTIGSIYIFG-TSELPSLWYSPPTRLSDLTALNOVKTDIEASTDNGTNT---G 564
Db 615 LQFTTKRTNKYALTNNFEGLINISGKVNISMVLPKNEGDKRGTMYMLTSLANSESG 674
QY 585 TTTTATSSGTCAGT-----GNTNTSQVYNSNTLNTYRSFGIDSKPTSKANKIDETN 637
Db 675 ERLTITDSRGSDSAGLTOPYNNGISFNKDTTFENRARNVFDIKA-PIGINKYSSLN 733
QY 638 WADPNYIARVAYEVLGIONEIPITNAGNFIRNTIGVGFTSGSRVLRASYNDQRP 697
Db 734 YASFN-----GNISVSGGSVDFT-----LLASSNVQRP 763
QY 698 TGNFOPLVYFGYLGIOQTRTGTFWTGTYKLLNSPEYDVLDSRVGTETNQFRT----- 752
Db 764 G-----VINSKXYFNWSTGSSLRKFTSGSTKTFGSIE 795
QY 753 -SLTPVWGGLTEGARSFSNTPYIRAOGDPRESISQSGSDMTYETIOGVLPDG- 810
Db 796 KDLTLATGNTT--LLOVEGTDMGIGKIYAKKNITTEGG--NITFGSKRAVTELEG 850
QY 811 --IRNNLNVGVKASFLNSNR-----NPNGL-----EMIAA 840
Db 851 VTINNANVTLLIGSDPDNHQKPLTIKKDYIINSGNLTAGNIYINAGNLVESNANPKAI 910
QY 841 TTYLRQIGLARTSGLPNOOPFGTTHQVIVSPGD-QFSSIKN-----IRTI 886
Db 911 TMTFVVGGLFDKNGKNSN-----ISIAKGARFKDIDSKMLSTITNSSSTYRTI 960
QY 887 FPGNOIMYFLFTNEN-----NKSSVYTLRL----- 911
Db 961 ISGN-----ITKNKNDLNTNNGSDTEMOIGDVSQKEGNLTISSDKINITKOITKAG 1014
QY 912 ADSNPDASSFSPTSLIDVNEIGVILLPLDNSF--YTVNAGVNLFPSSNPGSPSY 967
Db 1015 VOENSDSDATNANLTIKTELKLTODLINSGFNAEITAKGSDLTIENTN-SADGTN 1073
QY 968 TAVNTFNQNLSDIAFEGSGAKYTSDFWGTIOFKRPDEYLIQNGFTSOVARNFVTNQSFLNS 1027
Db 1074 AKKVTFNO-VKDSKISADGHKVT-----LHS 1098
QY 1028 LVDFETPANGTNRVVVDDGNLTNOLPLKVOIOYLDGKYDAK--LKNNNLVTPSYNN 1085
Db 1099 KVTSSGNNTT-----DSSDNAGLTI-----DAKNVTYNNNTTSHKAVS 1139
QY 1086 FGLPSPWVP---TAIGSTLGILAIMIILGLAIGIPLRQORKLQDGFKTFKVDITLA 1142
Db 1140 ISATSEIITTKTGTINATGNEIATQNGSIIG-----GIESSSQSV-TLTA 1186
QY 1143 -----AVGSYKKIITOTAN-----YKKRAALGAGKSGD 1172
Db 1187 TEGALAVSNISGNTVTVTANSALTTLAGSTIGTESVTTSSQSGD 1232

```

```

RESULT 13
US-10-011-366-6
; Sequence 6, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICIL-
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NO. US20030054493A1-2001
; CLASSIFICATION: <unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 3.08; Score 190; DB 15; Length 2710;
Best Local Similarity 19.38; Pred. No. 0.0002;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;

QY 42 VNQARTLDANSVRLAGISGLSNFVRLRVDDNFTFANGTIIKLDSFTKPL----- 94
Db 1071 INNSLSIAATVAVSIGAEVITFLPLIAGISGIFSLVNNELIHDKATSVNVPNHL 1130
QY 95 ---YGLDSDDCGGKRYKQIVSDYTTSSNRFDQRTFAYVALVNDANVHLKRI 149
Db 1131 ESKRYGPKLTREDP---KILVPIDDLVISEIDF-----NNN 1162
QY 150 SNRIQNRN-----NNSKPYIGGVDPNPAHRIRTDGTCKNFNFNQGEIVNDFIDAPI 203
Db 1163 SIKLGICNIIAMEGSGCHVTYGNIDH-----FESSPSISSHIPSLSIYSA-- 1207
QY 204 LPKDLHPDWNYLTIORKI--LPNDVNTAVVPWPVGRV-----SGTNADDMGFCGNG 253

```

Wed Oct 15 11:17:05 2003

us-09-901-572a-4.rapb

Page 12

```

Db      1208  -----IGIETENLDESKITMMLPN-APSRYWMEHGANPGLRSLJENDGTRILLSIRDLVPG 1262
      254  Q-----ITNDPIAO-----TKTTTNOQNSTPNSGAMG 283
      1263  KETWREYAFEDVAITLTKPYVEDNINIKIKLDDKDRNEMPTITTEIRKSLT---SFGD 1319
      284  ANNRRDSQNLNKHRIKTSFOLDERINTNSRIGNRNNNSKEVIGGVDPNPAHVIREFDDGT 343
      1320  AGGYLSLL-----STPISTNINLSKDDLM-----IPININEVEREISIEGT 1363
      344  KEFETNOTOGEIYNDFILDAPIPLPKDLHPDMNLTIOKILPNVNTAVNAPWPGYRSGT 403
      1364  K-----KGLIKD-----VLSKI---DINKNL-----LIGN 1387
      404  NADGMPDGNQOITNDPIAOCTTTTNDQNPSTFN-----SGAMPGANNRYDSOL- 454
      1388  QTD-----PSGIDNKKRIPLTCBLDDKISLIEINLVAKSYSLISGDKNYLISNLS 1442
      455  NVKHRIKTSFOLDEKTYPEWGTGSEENKNTIRLALGSLPSNERWYILDPGPVOLTLED 514
      1443  NTEIKINT-LGIDSKNIAVNYTDESNNKYFGALSKTSOKSIH-----KKD 1488
      515  SVNFSRKLXNSV---NSLSFGDSIYIFGTSPLSL---WYSSEPTLS-----DLAL 563
      1489  SKNLT-EPYNDSTLEFNSKDFIADIVFPMKODINTITKGYVDNNTOKSIDFSISLVSK 1547
      564  NOVKTDDIEASSTNGTTNGTTTADTSSTGSGAGGNTNTSQTSPNTINTYRSFGI 623
      1548  NOVU-----NGYLINESVSYSLDFEYKNSDGHNTSNMNFLO-NISFWKLEGF 1597
      624  DSKPTSAK-----IDETNMADPNYEARIYAEXYRLGIONPIPTMAGNFRNTIGVGFT 679
      1598  ENINNYIDKTYFLVKTNLG-----YEFICDNKNKIDI-YEGEMWTSKSTIIS 1647
      680  STGSRVYLASTNGOQRPFG-----NFOPLYEYGYGQOQRTGTF----- 721
      1648  GGNRNVYVPIYNPDP---TGEDISTLSDFSTYPLKIDRYINKVLIAPDLYTSLININTN 1704
      722  -----WGTYKLLNNSPYDV-----LDSF---RVGTETNOFRRTSLTPYMGVLYEEG 767
      1705  YSNEYYPPEIIVNPTEPFHKKNVINILDSSEFEYKWTGSDP-----ILVXYLEE- 1754
      768  ARSFSNTPYTRAGDTPEBSRSTFQSGYSDNTYEYIOSVYLGFGDGINNNLVNGVAKSFLNS 827
      1755  ---SNKKIL-----OKIRKILSNTQSPFKMSIDFEDIK-KLSLGYIMSNKSF 1800
      828  NRPNP-----NGLEMIATTYLRSQ-----IGIARTSGL-----PNOQPFQTHQVIS-- 870
      1801  NSEKELDRDLHGFKLIDNKTYYYDEDSKLVKGLININSLEYFDPILERNLVGWQTINGK 1860
      871  -----VSPGQOFSIKINIFTIPGNOLMYFLFTENNKKSSVYTLRLADSSNPDASSSP 925
      1861  KYEDINTGAALTSYK-----LINK---HFYNNNG-----VMOGVYKRGDGEYEPAP 1907
      926  TSLIDVNEIGVILLDNSEYTVAGANVALFSSNCGSPGSAVMT-----FNOLISDIA 981
      1908  ANTONNIEGOAI-VYOSKFLTLN--GKYYFEDNSKAVTGKAILINNEKYYNP-NATA 1963
      982  FEG---SGAKY-----TSDPMGTIO-----FKPDEYLIONGFTSOVAREFTVNO-- 1022
      1964  AVGLOVIDNNKIYENPDALISKGMQYNGSRYYEDDFAIAFGKTYIDGKHFYDSDOC 2023
      1023  -----SFLNSLVDETPANAGTNYRVVDPDGULTONLPLKYOIOLYLGK---YYDAKU 1073
      2024  VKIIGVSTSGEYEFAPANTYNN-----NEGAIYQSKFLTLNGKKYYPD--- 2071
      1074  KNNNLYT-----PSYNNFGALPSM 1092
      2072  NNSKAVTGLQITDSKKRYFNNTTARAALGW 2101
```

RESULT 14

```

US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276a1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRN
; ORGANISM: Haemophilus influenzae
US-09-797-862-33
Query Match      3.0%; Score 189.5; DB 10; Length 2353;
Best Local Similarity 18.6%; Pred. No. 0.00017;
Matches 256; Conservative 177; Mismatches 479; Indels 463; Gaps 67;
30 KOSDKSNDTQLVNQAFLDANSVRLAGLGNGSLFMTVLADVDNFTANGTITIKLDS 89
1141 KQSEK--DFYSLADLT-GLSTILGGANGRNDTGTVI-NKGGILITLNGAAAGTDA 1196
90 FTRPLVIGLSDDCGGYK---VKQIVSDYTSRRNRPQROTQRAYVALLV----- 136
1197 SNGNTISVTYDGIAGNKEITVKSALKTKYKDTQNTAETQDKEHAAVNAANEVEYVG 1256
137 -----DEANVHLKRIINTSNRI----- 153
1257 NGATVSAKTDNNGKHTVTDVAEAKYDGLKEDTQKIKLKVNTDGNLTLVDTAKGAS 1316
154 -----GNRNNNSKRVYIGGVDPNPAHVIREFDDCTKRFNTOTOGELYN 195
1317 VAKGEFNAVTTDATTAQGTANERGVVYV-----SNGATATFEDKKKVAIVG 1365
196 DE--TIDAPILPRDLHPD-----WYNLYQK----- 220
1366 DVAKAINDAAAFVKEVNDGATIDDSPTDGDANDALKAGDPLTLKAGKMLKYKROGKNT 1425
221 -TLPMND--VNTAVVPWVY-----GRVSGTNADDGM---PDGNGQITN--TDPITAOQRT 266
1426 FALANDLSVKSATVSDKLSLGTNGKKNVITSDTKGLFPAKDSKTGDDANHILNGIASLT 1485
267 TTDNONPSTFNSGAPGANNRRYDSOLNKHRIKTSFOLDERINTNSRIGNRNNNSK--- 323
1486 DTLNSGATNTNIGSGKIDTNEKKAASAKVDLNGW-----NVRGVKPAASANNQVENID 1539
324 -----FVIGGVDPNPAHVIREFDDCTKRFNTQNGELVNFILADAPILPKDLHPDW 374
1540 FVATYDTPVDSGDKDTSTVTVESKDNCKRFEVKTIGAKTSYINK----- 1583
375 YN--LYIQKILLENVDNFTAVVPPVYGRVSGTNADDGMPDCGNGOITNDPIA-----Q 425
1584 HNGKLFEGKELKADANNNGVYVTEETDK-----DEGNGITAJAVADAIVNAKAGWR 1632
426 TKTTTND-QNPS-----TF--NSGAMPGANNRYDSOLNKHRIKTS--FOLDEKF 470
1633 VKTIGANGQNDDEFATVAGTNYVFEADNGTIAEVTKANGSITVYKYNKVAVDGLKLDGDK 1692
471 VYPEWGTSEENKNTIRLATGSLPS-----NERWYILDI--PGTQV 509
1693 IVADTT-----VLTVAQGVTAJAPNNGDCKKRVASGLADALNKLSWATAGKGTGEV 1745
510 TLKESVNVFSLRYLNSVNSLSF-IGDSIYIFGTSPLSIWYSSEPTTRISDLTALNOYKT 568
```

Db 1746 ----DPANSAGO-EVKAGDKVTEFKADNLIKI---KOSGKDFYSLKKEIKDLNLS----- 1791
QY 569 DDEASSTNGT-----TTNGTWTADTSSGSTAGTGNNTTSSO----- 608
Db 1792 --VEFDANGTGSESTKITKDBGLTTPPANGAGACAGANTANTISVTKDGISACNKAATNV 1849
QY 609 -----TVSNPTL-----NTRYSF-----GIDSKPTSAKIDET----- 636
Db 1850 VSGLKFGDGHLLANGTVADFEKHVDNAKDLNLNDEKGDANNPTAADMTAAVGLDRGL 1909
QY 637 NW-----ADPNTEARITAEYRLGICQ-NEIPTNAGNPIRNTIGVGFTSGSRV 686
Db 1910 GWVYASDKTTGEPN-----QENYNAOVNANNEVKFKSG-----NGINVSAGKTLNCTRY 1957
QY 687 LRASVNGDQRPNGNPFPLVPGYIGYQQ--TRTGFWYGYTKLLNNSPYDVLDSRVTG 744
Db 1958 TEELAKGEVYKSNFT-----VKNADSETNLVKVGMY-----SKEDIDPATSKPMVG 2007
QY 745 ETNQPR-----RTSLTYPYMG-GYLTF-EEGARSFNTPYIRAQDTPESRSIF 790
Db 2008 KTEKYEVENGKTVSANGSKTEVTLTNKKSGYVTCNGVADALAKSGFELGLADAELAKAF 2067
QY 791 QSGYSN--TYEYIOSVLCFGDIR--NNLVGVKASSFLNSRPNGLMIAATYLS 846
Db 2068 AESAKDKOLSKDAEYVNAHDKVRFANGINTKVSATVEST--DANGDKV--TTTFVKT 2122
QY 847 QIGLARTSGLPNQPGTGT---HQVTSVBCDQFSSIKNRTITFPNQMLYFLETENN 903
Db 2123 DV-----ELPLTQYNTDANGKIKYKADGK-----WYELNAGGTAS 2159
QY 904 SSVYTLRLADSSNPDASSFSPSLIDVNEIGVILFLDLSFTYVNAAGVAFSSNPGS 963
Db 2160 NKEVTL-----GVVDANGK---KVKVTENGA-----DKWYTT-NADG----- 2193
QY 964 PGSYAVNTPFNQNLSDIAIEGSGAKYTSDFWGTIOKRPDE-----YLIONGTSOYARN 1017
Db 2194 ---AADKTKG-----EVSNDKYSTDEKHVYRLDPNNQSGKGVVIDNVANGELS-- 2239
QY 1018 FYTNQSFNLSDVDFTPANAGTNYRVVDDGNLTNQLPLKVOIQYLDCKYVYAKIKNN 1077
Db 2240 -ATSTAIINGSQLYAAKGVTN-----LAGOVNNLEGKVNKVG--KRAD 2280
QY 1078 LVTFSYNNFGALPSWVYPTAIGSTLGLIAMIIL---GLAIGIPLEAORKLQDKG 1129
Db 2281 AGTASALASQLPQATMP---GKSMVAIAGSSVYGQNGLAIGV-----SKISDNG 2327

RESULT 15
US-10-092-880-2
; Sequence 2, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE: HAEMOPHILUS
; CURRENT APPLICATION NUMBER: US/10/092,880
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-2

Query Match 2.9%; Score 189; DB 14; Length 1536;

Best Local Similarity 19.5%; Pred. No. 9, 8e-05;
Matches 266; Conservative 157; Mismatches 488; Indels 452; Gaps 67;
QY 27 PEGKQSD-----KSNNTQLVNAOKRLDANSV-RLAG-LGQNGSLFNTVLRVDDNPTRA 79
Db 105 FNIDQENWOFLOENNNSAVFNR--VTSNQISQLGIILDSNQVFL-----INPGIT 156
QY 80 ANGTIKLDSFPPVGLGLSDDCGGYKKKOIYSDTTSRN--RFDOROTRA-----YYA 132
Db 157 GKDALINTNGFTAST--LDISNE---NIK-----ARNTFEDTKDAELAEIYNHGL 202
QY 133 LAYDEANVHLKRIINTNSNRIGNRNNNS-----KFVIGVDNPA-----HVIR 175
Db 203 ITYVKDGSVNLIGKVKNEGVLISVNGSISLAGQKITISDINPTTYSIAAPENEAVN 262
QY 176 FTDDGTFKFNFTNOTGCEIYNDFLDAPILPKDLHPWYNLYIORKILPNDVNTAVPWPV 235
Db 263 LGDIFAKGNINVRATIRNOGKLSDSVSKSG--NIVLSAKGEAEIGCVI----- 314
QY 236 GRVSGTNADDGMDCGNGOITNDPIAOKTKTTD---NONPSTFNSGAMPANRRYDSOL 292
Db 315 -----SAONOAKGKMLITGDKVTLKTGAVIDLSGKEGETY-----LGDEREGKN 363
QY 293 NVKHKRTSFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHIREFTDGTKEFNFTNQ 352
Db 364 GIOLAKKTSLKSGSTINVSKEKEGR---AIWMDI-----ALIDG--NINAQS 408
QY 353 GEIYND-----FLIDAPILPKDLHPWYNLYIORKILPNDVNTAVPWPVGRVS 401
Db 409 GDIAKTGGEVETSGHDLFIKDNAIVDA--KEW-----LLDFD-NVSIINETAGR-S 455
QY 402 GTFNADGMDFCGSGQIT---NTDPIAOKTKTTDNONPSTFNSGAMP--ANNR--VD 454
Db 456 NTSDEDEYSGSGASSTPKRNKKTTLTNLTLE---SLUKKTEVNTANQRIYVNSI 511
QY 455 NVKHKRTSFOLDEREVEYDEMT-----GSEENKNITRLATG-----SLPSNRYMI-- 500
Db 512 NLSNGSLTL-----WSEGRSGGGEIINDIT---TGDDTGANLTIYSGGMDV 557
QY 501 ----LDITGTPQVYAKED-----SVNV-----FSRLYLVNS-LSF 532
Db 558 HKNISLGAOGNINITAKODIAFEKGSNOVITGCGTITSGNOKFRNNNSLNTGSGLOF 617
QY 533 IGDIVIFG-TSELSLWYSEFPTRLSDLTALNOVTDLDIASSTNGTGTN---GTT 587
Db 618 TTKRTNKYALTNKFEGTLNISGKVNISWVLPKESGIDKFKGTYNNLISLVNSEGEFN 677
QY 588 TADTSSGSTAGT-----GNTTNTSQTIVSNPTLNTYRSFGIDSKPTSAKIDETNMAD 640
Db 678 LTIDSRGSDSAGTLTPYNLNGISFPMKDTTFENRNARVAFDIKA-PIGINKYSLNVAS 736
QY 641 PNVEIARITAEYRLGIONEIPITNAGNFIIRNTIGVGFTSGSRVYLRASVNGDQRP 700
Db 737 FN-----GNISVSGGSVDFT-----LLASSNVQYPPG-- 764
QY 701 FQFLVYVPGYLGQYQRTGTGFWYGYTKLLNNSPYDVLDSRVTETNOFPRT-----SL 754
Db 765 -----VINSKFFNVSTGSLFKFTSGTKTFSIEKDL 798
QY 755 TYPVMGYLTREGARFSNTPYIRAQDTPESRSIFQSGYSDNTYEYIOSVLCFGD--I 811
Db 799 TLNATGNGIT--LLQVEGIDGMIGKIVAKNITEGEG--NITFSSRAVTEIEGNVT 853
QY 812 RNVLNVGVKASSFLNSRNP-----NPNGL-----EMIAATY 843
Db 854 NNNANVTLLIGSDDNHOKPLTIKKDVIINSGLNLTAGCIVNIAGNLTVESANAFKAITNF 913
QY 844 LRQIGLARTSGLPNQPGTGTGHQVTSVSPGD-QFSSIKN-----IRTIFFP 889
Db 914 TENVGGLFDNKGKSN-----ISIAKGARFADIDNSKSLSTTNSSTYRKIIISG 963
QY 890 NQMLYFLFTNEN-----NKSYYTLRL-----ADS 914

us-09-901-572a-4.rapb

Wed Oct 15 11:17:05 2003

```

Db      964 N-----ITKNGDLNITNEGSDTENQIGDVSOKEGNLTISDKINTKQITIKAGVDG 1017
OY      915 SNPDASSSEFPTSLIDVNEIGVILPLIDNSF-----YTVNAGNVALFSSNPGSPGYTAV 970
Db      1018 ENSDSDATNNANLTIKTELKLTQDLNISGFNKAEITARKDGDLTIGNTN--SADGTNKK 1076
OY      971 NTFNOLSDIAFEESGANVTSDFWGTIOFKPDEYLIONGFPTSQVARNFVTNOSFTLSLYD 1030
Db      1077 VTFNQ-VKDSKISADGHKV-----LHskVE 1101
OY      1031 FTNANAGTNRVVDEPDGNLTNQLPLKVQIQYLDKYYDAK--LKNNILTFPSYNNFGA 1088
Db      1102 TSGSNNTF-----DSDNNAGLTI-----DAKNVTYNNNTSHKAVSISA 1142
OY      1089 LPSWVP--TAIGSTLGLIAMITLGLAIGIFLRAORLQDKGFKTFKKVDLTIA--- 1142
Db      1143 TSGEITTKGTITNATGNVEITAOGSILG-----GISSSGSV-TLIRTEG 1189
OY      1143 --AVGSVYKKIITOTAN-----YKKPAPALGAGKSGD 1172
Db      1190 ALAVSNISGNTVTYANSALITLAGSTIKGTESVTTSSQSGD 1232

```

Search completed: October 10, 2003, 16:45:49
 Job time : 281 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 16:32:21 : Search time 23 Seconds
(without alignments)
2251.673 Million cell updates/sec

Title: US-09-901-572a-4

Perfect score: 6413

Sequence: 1 MNISKKLKSYTLIGLAVG.....APTPTAPKPAKPAKPAKPE 1224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328	5.1	1627	1 US-07-665-792E-9	Sequence 9, App11
2	252	3.9	10162	4 US-09-134-001C-3159	Sequence 3159, Ap
3	224.5	3.5	2504	4 US-09-328-352-5821	Sequence 5821, Ap
4	205	3.2	2123	3 US-08-968-685A-10	Sequence 10, App1
5	199	3.1	1612	1 US-08-169-927-2	Sequence 2, App11
6	192	3.0	1536	4 US-09-206-942-67	Sequence 67, App1
7	190	3.0	2710	1 US-08-480-604A-6	Sequence 6, App11
8	190	3.0	2710	2 US-08-405-496A-6	Sequence 6, App11
9	190	3.0	2710	3 US-08-915-136-6	Sequence 6, App11
10	190	3.0	2710	4 US-08-957-310-6	Sequence 6, App11
11	190	3.0	2710	4 US-10-011-366-6	Sequence 6, App11
12	189.5	3.0	2353	3 US-09-377-155-33	Sequence 33, App1
13	189.5	3.0	2353	3 US-08-913-942-4	Sequence 4, App11
14	189.5	3.0	2353	4 US-09-669-974-33	Sequence 33, App1
15	189.5	3.0	2736	4 US-09-252-991A-30227	Sequence 30227, A
16	189	2.9	1536	1 US-08-038-682-2	Sequence 2, App11
17	189	2.9	1536	2 US-08-302-832-2	Sequence 2, App11
18	189	2.9	1536	2 US-08-530-198-2	Sequence 2, App11
19	189	2.9	1536	2 US-08-469-880-2	Sequence 2, App11
20	189	2.9	1536	2 US-08-728-470-2	Sequence 2, App11
21	189	2.9	1536	2 US-08-617-697-2	Sequence 2, App11
22	189	2.9	1536	3 US-08-719-641-2	Sequence 2, App11
23	188	2.9	2411	4 US-09-268-347-36	Sequence 36, App1
24	187	2.9	1529	3 US-08-728-470-10	Sequence 10, App1
25	187	2.9	1529	3 US-08-719-641-10	Sequence 10, App1
26	185	2.9	2354	4 US-09-268-347-47	Sequence 47, App1
27	183	2.9	1600	2 US-08-617-697-10	Sequence 10, App1

28	182.5	2.8	2314	4 US-09-268-347-49	Sequence 49, App1
29	180.5	2.8	2048	4 US-09-268-347-48	Sequence 48, App1
30	178.5	2.8	804	4 US-09-328-352-5545	Sequence 5545, Ap
31	177	2.8	1833	4 US-08-621-944A-4	Sequence 4, App11
32	177	2.8	1833	4 US-08-945-567D-4	Sequence 4, App11
33	177	2.8	1922	4 US-08-621-944A-3	Sequence 3, App11
34	177	2.8	1992	4 US-08-945-567D-3	Sequence 3, App11
35	175	2.7	2057	4 US-09-499-903-2	Sequence 2, App11
36	174.5	2.7	1278	4 US-09-604-957-3	Sequence 3, App11
37	173.5	2.7	1848	3 US-08-296-791-6	Sequence 6, App11
38	173.5	2.7	1848	5 PCT-US95-10661A-6	Sequence 6, App11
39	172.5	2.7	1287	1 US-08-200-232-2	Sequence 2, App11
40	172.5	2.7	1287	5 PCT-US95-02219-2	Sequence 2, App11
41	172.5	2.7	1287	5 PCT-US95-02219A-2	Sequence 2, App11
42	172.5	2.7	1335	4 US-09-134-001C-3716	Sequence 3716, Ap
43	172	2.7	1231	4 US-09-071-035-420	Sequence 420, App
44	172	2.7	1265	4 US-09-071-035-418	Sequence 418, App
45	171.5	2.7	1073	4 US-09-206-942-49	Sequence 49, App1

ALIGNMENTS

RESULT 1
US-07-665-792E-9
; Sequence 9, Application US/07665792E
; Patent No. 5281694
; GENERAL INFORMATION:
; APPLICANT: Baseman, Joel B.
; APPLICANT: Su, C. J.
; APPLICANT: Daillo, S. F.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: PRODUCTION OF MYOPLASMAL ADHESINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5-1/4" DSDD Diskette
; COMPUTER: Compaq (IBM PC Compatible)
; OPERATING SYSTEM: MS DOS 3.31
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/665, 792E
; FILING DATE: 19910607
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/118,967
; FILING DATE: NO. 5281694. 10, 1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/588,886
; FILING DATE: July 27, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/616,111
; FILING DATE: NO. 5281694. 21, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/697,349
; FILING DATE: May 8, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Melinda Patterson
; REGISTRATION NUMBER: 33,062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1592
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1627 Amino Acids
; TYPE: AMINO ACID
; TOPOLOGY: Linear

QY 209 HPDMVLYTOR-----KILPND-VNTAVVMPVGR-----VSG-----T 241
Db 1524 -----YIEKLGONLSNTNNVINYFVSDKVNPSIT---YGNVHHYVSEETKNTI 1573
QY 242 NADDFGDCNGOITTDPIAOTKTTTDN---QNPSTNS-----GAMPANNRD 289
Db 1574 NVND---NYGLNVTASTDSALITMTNRNNELVGAQAVNJUSINKIKVKATKDSGNESTI 1630
QY 290 S-----OLNVKRIKTSFOLDERINTNSRIGNRNNNSKFVGVGDNPANHVRFDND-G 342
Db 1631 SFTVNIKPLNEKRIITSSS-----NOTPRISSIONANLSTEDONRVKSSLSMKIILG 1685
QY 343 TKEFNTOGELIVNDFILDAPILPKDLHPDMVNLXIQRIKLPNDVNTAVVMPVGRVSG 402
Db 1686 TR-NYVNESN-----NDVRSQVVS-KVNR-SG 1709
QY 403 TNADDFGDCNGOITNT-----DPIAOTKTTTDNQPSTNSGAMPANNRDYS 452
Db 1710 NNATVNTTTFSDGTNTITVVPKVHVLLEVVPTRTTRVGOQ---FPTGKSTPNDFFS- 1765
QY 453 QLVNKHRIKTSFOLDERKFPVPEWGTSEEN-----KNITRLA 488
Db 1766 -----LRTGGVVDARIYVWVNNOGPDINSNOIGRDLTLAEIFFPOGETTPIRKDTIYL 1818
QY 489 TGSLSPSN-----ERYW--ILDI-----PGTPO----- 508
Db 1819 SOSIPKOIYETTINGRFNSGDAYPGNFVQAVNQWPEHMDPFMAQCSGTPSSRNAGSFT 1878
QY 509 --VTL-----KESDVVPSRLIYNS--VNSLSFI-----GDSITIFGTSELSLWYTF 553
Db 1879 KRTVVVYONGOTENVVLEKVKPNKPVJDSNSVYSKQOLNQQOILVRNVPOMQAVLYXO- 1937
QY 554 PRTSLDLALNQKTDIDEASSTDNGTTNGTTTADTSSGSP-----GAGSTNTNTSQ 608
Db 1938 -----SNGTIVPNTNTTIDSNGIATVTTIOGILPNTAKTIS 1974
QY 609 TVSNPTLNTYRSFGIDSKPT-----SANKIDET-----NMADPN 642
Db 1975 MTNNVYTTKONSGIASNTEDISVSENSDQVNTAGMAKNDGIKIIGTNYNFDEN 2034
QY 643 VIEARIYAEYRLIONE-----IPIT-----NAGNF 668
Db 2035 SFISNIPAHSTLTWNEEPSNMKNIGTTTCTVTLPHOGTTRVDPIPIIYFVYAKNP 2094
QY 669 IR-----NTIGVGFTSGSRVYLRSYNDQDQPTGNF---QP 703
Db 2095 VRQOKGBRLNTGDIYNTIIFENNRRLGGT-----ASWMDNROPDKNIAGVON 2142
QY 704 FLYVFGYLG-----YOOTRTGT-----WYGTYKILNNS---PY 734
Db 2143 LIALVNVPGISTPLEVPVKVWVYNFDETOPRIKIQODTTPPKGTMACYKHLNGBEPLI 2202
QY 735 D-----VLDSFRVGTETNO-----FRRTSL-----TYPM-----GGYLTREGAR---SF 771
Db 2203 DGMKFYWNQOSTGTSTDOQSLAYTRTPEVKTGTIVDVVNSMNGVMQTSQSAKFIYTNAK 2262
QY 772 SNRPYIRAO---GD---TPES-RSIFOSGYDNTYEVYQSVLGFDCIRNNLNVGKASSFL 825
Db 2263 PNDPTTQOSKGTGVYTPGAVRNILISG---TNDYIOA---SADKIYINRN--GNKLITTFV 2315
QY 826 NSN-----RPNNGLEMIATTYLRSOIGLARTSGLPNOOPFGTTHQVIVSVPDQ 876
Db 2316 KNNDKRWYETGSPDINGI---GPTNNGTAISLR-----LAVRGRD- 2354
QY 877 FSSIKINRTIFPGNQLMYLELTENNKSYYTLRLADSN-----PDAS-S 921
Db 2355 --SIEAIATGSGEFTI-----STASISEIYIKAPQEQVATHYDNGTFDILLPDSRN 2406
QY 922 SFSPTSLIDVNEIGVILLPLDNSFYVNAAGNVA--LFSNPSGPGSYTAVN---TFN 974
Db 2407 SLNPTREVEIN-----YTEKLNGETQOSFTITKNNNGKWTNNKPNVYEFN 2453
QY 975 QNLSDIAFEGSGAKYTSDFWGTIQFRPDEYLIONGFTSQVARNFVTNOSFLNSLVDPPTA 1034

Db 2454 QDNKGVESANTIKPSQI--TTPRAGQNTEN-----TNPVIOA-----PA 2495
QY 1035 NAGTNRVRYVDPG-NLTNQNLEPLKVOIQYLDCKYTDAKKNNLNTYFSINNGLAPSW 1093
Db 2496 QHTLTINEIYKEGQVNTDNDINNVAOV-----PNKRVAIKOGN--ALPT-- 2539
QY 1094 VPTAIGSTGLIAIM 1108
Db 2540 -NLGGSTSHIPVYI 2553

RESULT 3
US-09-328-352-5821
; Sequence 5821, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5821
; LENGTH: 2504
; TYPE: PR
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5821

Query Match 3.5%; Score 224.5; DB 4; Length 2504;
Best Local Similarity 18.8%; Pred. No. 8.8e-07;
Matches 267; Conservative 179; Mismatches 442; Indels 529; Gaps 68;

QY 34 KSDNTQLVNQAARTLDANSVRLAG-----LGO-----NCSLENT-VLRVDNDFI- 77
Db 876 KTELNTKIDNTKTELEKKGINFAGNSGADVHRKIGDKLNTIVGAAASTPRAKTSGENIIT 935
QY 78 -TANGTII-----KIDSTK-----PLYGDLSD-----DGGYKVVQI--VSD 114
Db 936 RTTQDGIQIELKDSKFSVYTGNTLTNGTLTKESPSITKOGINAG--KQITNVAD 992
QY 115 YTSRNFQDQRTRAYVALLVNDENAVNLKRITNS-----NRIGNRN---NSK 161
Db 993 GINAKDAVNDQ-----LTQKKNLNGTITDINNOLNDAKKDLGQIADTKKLNNAK 1045
QY 162 FVIGGVDPNPAHVIRFTDDGTKEFNT-----NOTGELVNDF---ILDAP--LPR 206
Db 1046 KDLGD-----QIFDTNFKLNNTKDQLTTOINDFTTELNTIGNTKTELNSKIDSTKT 1097
QY 207 DLHPDMVNLX-----IQRI--LPNDVNTAVVMPVGRVSGTNA-----D 244
Db 1098 ELENNKGINFAGNSGNDVHRKLGKRLNIGGAAASTPRAKTSGENVITRTTQDGIQIELK 1157
QY 245 DGMFDCNGOITNTDPIAOTKTTTDNQPSTNSGAMPAN-----NRYDSOLNVK 295
Db 1158 DSKFD-----SVTGNLTTLNNTGLTIKESPSITQGIAGSKQITNVADGGINAADA-VNVD 1212
QY 296 HRIKTSFOLDERINTNSRIG-----NRN--NSKFIYGVDPNPAHVIRFTDD 341
Db 1213 QLVKVENLNGRITDTNNOINDAKKDLGQIADTKKLNDAKKDLG-----IQITDT 1264
QY 342 GTKEFNT-----NOTGELVYNDF---ILDAP--LPKDLHPWYLY-----IQ 380
Db 1265 NTKLNTTKDQLTTOINDTKTELNNJTGNTELSKIDNTKTELENNKINFAGNSGADVH 1324
QY 381 RKT--LPNDVNTAVVMPVGRVSGTN-----ADGMPD----- 411
Db 1325 RKLGDKLNIVGGAASIPRAKTSGENIITTTQDGIQIELKSKSPDSVYTGNTTLNTNG 1384
QY 412 -----CGNGOITNT-----DPIAOTKTTTDNO----- 433
Db 1385 LTIKEGPSITKEGINAGSKQITNVADGINAKDAVNKSQOLDNLAAQNATDAAVVKYDAK 1444

434 -----NPS---T 437
1445 TDKDTLKKDQTVLDNVKAGHISSTSKAVNGSOLKHSINSIGNTVVPDGSIT 1504
438 FNSGAMPANRRYSOLANVH---RIKTSFOLDEFFVPEMTGSEENK-----I 484
1505 TNNIGCTGNININDAISEVKNATKATTVTEGDIVVKEIVYKNDGSGINTEVATRKDITL 1564
485 TRLATGSLPSNEKRYWLDIPGTPOVYLKEDSVNPSRLYLSVNSLSTGSIYIFGISE 544
1565 NSVTGDTYVLANNG--LTIKDGPSTI--KDGINAGSKOITTVADGIN--AKAVNDOLTK 1619
545 LPSLWYYSFPTRLSDL--TALNOYKTD-----DIEASDTNGTGTNGT 586
1620 VKE-----NLNGRITDNNLNDAKDKLGNOLADTNKNLNDAKKDLGNQITDNTLNTK 1675
587 TTADTSSGSGAGCTNTSOTVSNPTLNTYRSFSGIDSKPTSAKIDET-----N 637
1676 DQLTTOINDTKTELNTIGTKTELN-----SKIDSTKTELENKGIN 1717
638 WADPNVIEARLYAEVRLGIONEIPITNAGNFR-----NTIGVGFT-----STGSRV 685
1718 FAG-----NSGNDVHRKLGKEKINLTIGGAASTPAKISGENV 1754
686 VLRSYNGDQRPNTQPFVYFGLGYOQTRGTGTEWYGYKLLNNSPYDVLDSRVGTE 745
1755 ITRTODGIG-----IELKDSKEDSVTGTGTTLN 1784
746 TNOFRRTSLTYPVNGVYLTEEG-----ARSFSNTPYIRAGDTPES 786
1785 TN-----GLTIK-EGPSVTEKGINAGGKITTVADGVNAKDAYKSKDLNLAKONATD 1838
787 RSIFPGSYSDNTYEVYEQSLGFDG--IRNNLNWGVKASSFLNSNPNGLEMAATYLR 845
1839 AAV---KYDDAKTKRKYVLKKGDTVLDN---YKAGHISSTSKAVNGSOLKHSINSIK 1891
846 SOIGLARTSGLEPNQPGF--TTHOYVISTP---GDSSIKINITIPNGNOLWFLEPNE 900
1892 NSIG-----GNTVVPDGLSTNNIGTGKNININDAISEVKNAT-----1931
901 NKKSSVY---TLRLADSNPDASSFSPTSLID--VNEIGVILPLDLSFTYVNAAGNVA 955
1932 KAKTYTEGDINIVKETEYKNDGSGTNEYESTKDKLNLNSVTGIDVYLNNGGTLIDGSI- 1990
956 LFSSNPGSPGTYAVTFENONISDIAFESGAKIYSDPWGTIQFPEYLLONGFTSQVA 1015
1991-----TKDGVNAGGKITTDVA-----NGVIAQNS 2014
1016 RNFVNTOSFLNSLVDTPANAGTNYRVVDPDGNLTNOLPLKVOIQYLDG--KYIDAKLK 1074
2015 KDAV--NGAOVYHHSINSIKNSIGN--TYVNPDSGLTTNIGCTGNININDAISKVDEKVT 2071
1075 NN-NLVTFSYNNGA-----LPSWVYPTAIGSTIGL 1105
2072 NGVNDLTOKGLNFGANDOKTTOGKAVHRKLGDTITIV 2108

RESULT 4
US-08-968-685A-10
Sequence 10, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSLA, LAURA
TITLE OF INVENTION: MORAYELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968, 685A
FILING DATE: NO. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-864
TELEFAX: (212) 869-864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 3.2%; Score 205; DB 3; Length 2123;
Best Local Similarity 19.8%; Pred. No. 1.7e-05;
Matches 286; Conservative 157; Mismatches 502; Indels 498; Gaps 72;

24 SASFGKSDKSDNTOLVNOARTL-----DANSVRLAGGON-----GSLFNTVLSD 71
482 AATVG--OLKRVNPTAESALOTFTYKVDKNGANDANSKIITVYKNNKPGDGOVNT----- 535
72 VDDNFTIANGIITLIKIFRPLVGLDISDCGGRYKQVYSDYTSRRNFDDORARAY 131
536-----LKL-----KGNGVDVTTETNG-----TYTF 556
132 ALLVNDENVHLKRIINTSNRIGNNNNSKRVIG-----GVNDPAHVIR 175
557 GINONNGILVNSTLNNGLSVKNKNSKQIOVGADGITTFIDISNKPAGIEN---TFR 613
176 FTDDGTFKFTNOTOGELVNDILDA---PIPKDLHPDWNYLQKRIIPNDVFAVVP 232
614 ITRDGG--IGFANNVGS-----LDANKPRLPFTGINSKREL-----TVOQSAINP 656
233 WPGVAVSGTNADDGMPDCNGQIT--NTDP IAO TKTTTD-----NONPSTFNSGAMPANRR 287
657-----ATNG--GQIDFNNRUSTANTKSSGAITIKDLYNLNSQVPLTFAGDTGPNVTK 707
288 Y-----DSOLNVKHKIKTSFOLDERINTNSNRIGNR- 765
708 LGELIKYKGGKTTADDLTKNNIGVADSPDNLIVK--IAKTLSDLD--AVNTKTLTASDKV 765
319-----NNNSKF-----VIGGVNDPAHVIRFTD-----DGTREN 346
766 TVDSGNNTAKLONGDILTFSKONTGATPATNSKITTYVGD-----LKFIDNNGIALDGTTY- 820
347 FTNOTOGELVNDILD--APILPKDLHPDWNYLYL--QRIKLPNDVNT---AVVPMVAVGV 400
821 ITRKQVGFANQDGLDSKRPYLDK-----KIKVEVEITNGINAGGKATYGLSNTLT 874
401 SGTNADG-----MFDGNGQITVNTD--PIAQTKTITDNONPST 437
875 DATNATGHTVQLGIVSDTKTRAASIGDVLNMGFNLNKNGDAKDEVSITDYVD----- 928
438 FNSGAMPANRRYSOLANVHRIKTSFOLDEKFEYPMPTGSEENKNITRLATGSLPSNER 497
929 FINGNATTAKYTTDGKAS-----KVAVDVNVVDGTIHLTGADGNKNQIGKTYTTLKKT-- 981
498 TWIIDIPGTPOVTLK-----EDSVNVFSRLYLSVNSLSIFGDSIYIFGTELSLWYVSF 553

```

Db      982  ----DAKGDAINFVSNGDDKALINAKDIADNLMTLA--GEIRMTKGTAD----- 1026
OY      554  PIRLSUTLALNOVKRDDIDIASSTDNCTTNGTTTADTSGSGAGCTGTTNTSQTVSNP 613
Db      1027  -----TALQTFQVKVKV-----ENGDDN-----DADTIT-----VGDAKTNQ----- 1060
OY      614  TLNTYRSFGIDSKPTSAKIDETNMADPVIEARIAEYERLIGNEIPTNAGNFRTMT 673
Db      1061  -VNTLKLKGNKGLDIOFNR-----DQTVTFGITQYSGLRAGNNTTILNNGSLSTKRNA 1111
OY      674  G-----GVGFTSTGSRVYLRLASVNGDQRPNGFQFLVYFGYLGQYQRTGTG--FW 722
Db      1112  GNEQIOVGADGVKFAVNNGVV--GAGIDGTRITRDEIGFAGNGSLDKSKPLSLKDGIN 1170
OY      723  YGTYKLNNSPYDVLDSPRVGTETNFRRTSLTYPMGVL-----TE-----EGARSS 772
Db      1171  AGGKKTIN-----IQSGEIAONSND-----AVTGKRIYDLKTELENKISSTAKTQ 1216
OY      773  NPPYIRAGDTPESRSIFOSGYSDNTYEXIQSVLGF--DGIRNNLN-----VGKASSF 824
Db      1217  NSLHESVADDEGONNTVSNPYSYDTSKTSIDVITTAGENGITTKYKNKGVVRYGIDQTG 1276
OY      825  LMSNR---BNPNGLEMIATYTLRSQIGLARTSGLEPNQOPFGTHQVISVSPGDQSSIK 881
Db      1277  LTPKLTGVNNGNGKGV-----IDSQNGQNTITGLSN-----TLA 1311
OY      882  NRTTFPGNQLMFLFTNENKSSVYTLR-----LADSSNPDASSSPTSLDYNIEGY 937
Db      1312  NY-----TNDKGSVYRTTEQGIKIDEDKTRAA-----SIYDVLGAGFN 1349
OY      938  L-----PLLD--NSFYVNAAGNALFLSSNPGSPGSYTAVNTFNQNLSDIAFEGSGAK--- 988
Db      1350  LOGNGEAVLFPVSTYTDVNPADGNATYAKVYDDTSKTSKYVDVNVDDTTIEKKKLGAV 1409
OY      989  ----YTSDFWGTIOFKPDEYLQNGFT-----SQVANFYTNQSEFLNSLVDFTPANAG 1037
Db      1410  KTTTLTSTGTGA-----NKFALSNQATGALVKASDIAVHMLNTLSDIDQTAKAGASQANS 1464
OY      1038  TYRRVYVDDGNLTNQNLEKQIQYLDGKYDYDAK-----LKNNNLVYFSNNFG 1087
Db      1465  AGT---VDADGN-----KVYYDSTDNKYYQAKNGDGYDKTEKAKDKLVAQAQTPDG 1513
OY      1088  ALPSWVYPTAIGS-----TLGLAMITLL 1111
Db      1514  TLAQNMVKSVINKEQVNDANKKOGINEDNAFVKGLEKASDNKTKAATVGLNAV--- 1570
OY      1112  GLAIGTFLRAQRKLQ---DKGRTTFKKV--DTLTAAGSVYKKIITQ-----T 1155
Db      1571  -----AQTPLETFAGDTG--TTAKKLGELTIRKGGQDTNKLTDNNIGVAVAGTDFGT 1619
OY      1156  ANVKKKPAALGAGKSG---DKKPAALAPAPAKSAPKASSAPKPTGPK---SGAP 1207
Db      1620  VKLAKDLTNLSVNAAGTKIDKGVSVDSGQAKANTPVLISANGLIDGKVISNVGKGT 1679
OY      1208  KPT 1210
Db      1680  KDT 1682

```

```

RESULT 5
US-08-169-927-2
: Sequence 2, Application US/08169927
: Patent No. 5783441
: GENERAL INFORMATION:
: APPLICANT: Carl, Mitchell
: APPLICANT: Dobson, Michael E.
: APPLICANT: Ching, Wei Mei
: APPLICANT: Dasch, Gregory A
: TITLE OF INVENTION: Gene and Protein Applicable to the
: TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
: TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
: NUMBER OF SEQUENCES: 2

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Counsel, Naval Medical R & D Command
: STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
: CITY: Bethesda
: STATE: MD
: COUNTRY: USA
: ZIP: 20889-5606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/169,927
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/742,128
: FILING DATE: 08/09/91
: ATTORNEY/AGENT INFORMATION:
: NAME: Spevack, A. David
: REGISTRATION NUMBER: 24,743
: REFERENCE/DOCKET NUMBER: 75,976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-1022
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1612 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-169-927-2

```

```

Query Match      3.1%; Score 199; DB 1; Length 1612;
Best Local Similarity 17.5%; Pred. No. 3e-05;
Matches 228; Conservative 170; Mismatches 390; Indels 516; Gaps 59;

OY      14  GGLAV-----FGALGSAISFGFKOS-----DKSNDTQLYNQRARTIDANSVRLAGIG 59
Db      127  GCAAINANDLSGSGITFAVCPSEVLEFNILNPINSRSSHYLANSNKYV-----G 177
OY      60  QNGSLFNFTLRDDVDNFITAANGTIIKLSFTKPLKGLDSDCGGKYKQIVSDYTSR 119
Db      178  GNCIL-----NTTNFQIVSDNFTFAGITIN-----IDDCG----- 209
OY      120  NRPDQROTAYVALVN---DEANVHLKRINTNSNRIGNRNNSKRVIGVDNPAHVIRF 176
Db      210  -----LMENSTPDAAN-----TLNLQAGGWTIN-----FNGIDGTGLVLV 245
OY      177  TDDG--TKFNFTNOTGGEIVNDPILDAPILPKDLHPDWYNLYIQKILRPDVTAVVPP 234
Db      246  SKNGATERNVCTGCGNL-----KGI--ELMTAAVACK 278
OY      235  V-----GRVSG--TNADG----- 246
Db      279  LISLGAANAAYIGTDGAGRAAGFIYSVNGNATISGOVYAKNMVIOGASAGGYTFEH 338
OY      247  MEDCGNGQITNDPDLAQTCTTDNQNPSTFNSGAMPKANRRDSQLNVKHR--IKTSFOL 304
Db      339  IVDVGLGCTTNEFTADSKVILITENSNGSYNFGNL-----DTQILVDPDTKILKGNFTG 391
OY      305  DERINTNSNRIGNRNNSKRVIGVD-----NPAHV--IRFTD 341
Db      392  DVAKNNGNTAGVITFNANGALVASTDPNTIAVTINAIIEAGACVVELSGIHIDELGNG 451
OY      342  GTFENFTNOT-----QGEIVDEFLDA-----PIPKDLHPDWYNLYIQKILRP 386
Db      452  GSIFKLADGTVINGPYNQANLMMNNMLAAGSIDLDSALITIGDIGNGVNAALQHTTLAN 511
OY      387  DVNTAVVPMVGVGSGTNADCGHFCGNG--QITPTD-----PIAQK 427
Db      512  DA-SKILALDGANIIGANVGAIHFQANGGTIKLTGONNIVVFPDLIDTDTKTGVVDS 570

```

OY		428	T TTDNONPSTFNENSGAMPKANRRYSOLVNHKRIKTSPFOLDKEFYVPEDWGSEKKNTIRL	487	
Dd	: : :	:	:	: : :	606
OY	S L I N Q T F L I N G S I G V A N T K T L A D I N T -	-GSK-	--TIL		
OY	488	A T G S P E N E X Y R I L D I C H P O V T - - -	L K E D S V N F S R - - -	L Y N S V N S I S F G D S I Y I	539
Dd	607	N A G U A L A N E - - L V I E N N G S Y Q L M H N Y L I T K T I N A N Q O I I V A A D P I N T N T L A D G N T L			664
OY	540	E G T S E L P W Y S P E R L D J A L - - - -	-N O V T D I E A S T N G T T G T T Y T		588
Dd	665	-G S A E N D - L S T I H P A T K A N A D S I L N V K G V N L Y A N N I T T N D A N G S L H -	F R S G T S I Y		720
OY	589	A D T S S G S T G - - - - -	A C T - - - - G N T - N T S O T V S N F T L - - -	N T Y R S F G I D S K P S	629
Dd	721	S G T V G G G G H K L N L I I D N G T Y K F L D T T F E N G G T K I E G K S I L O I S N N Y T T D H V E S A D N T			780
OY	630	A N K I D E T M A D P N V I E A R I A E Y L G I O N E P I T N G N F I B N I I G G V F T S G S R Y V L A			689
Dd	781	G T - L E F V N - T D P I F Y T L N K O G A Y - F G V A K O V I I S G P E N I V F N E I G N V G - - - -	I V H G I		830
OY	690	S Y N G Q G R P T G N O P L Y F G Y L G Q O T R G T F W Y G T K L N N S P Y D V L D S P R V G T E N O F			749
Dd	831	M A N S - - - - -	I S F E N A S I G S T L F - - -	L P S G T P L D V I L - - - -	858
OY	750	R R T S I L T Y P W G Y L T E G A R S F S N T P Y I R A G D O P E S R S I P Q S G S D N T E Y I O S V L G F D			809
Dd	859	- - - - -	I K S F V G - N		865
OY	810	G I R N N L N V G K A S F L N S R P N G L E M I A T T Y L R S O I G L A R T S I G L P N O P F C T T H O V I			869
Dd	866	G Y T D N N N A I I V A V G I D S W - - - -	I N N G O I I D K K N I I		898
OY	870	S V S P G Q F S S I K N R I T F P G N O L M Y F L F E N N K S S Y T L R L A D S N D - - - -	A S S F S		924
Dd	899	A L S I G S D N S I T V A N T L Y S G - - - -	I R T K N Q G T V - T L S G M N N G T I Y G L G L E N G S		951
OY	925	P - - - - T S L I D V E I G V I L - - - -	P L D N S F Y T - - - - -	V A A G V A L F S	958
Dd	952	P K L Q Y T F T D Y N L N S I T A N N V T I D D V T L T T G G I A G D F D P A K T I L G S V N A N A V R E D			1011
OY	959	S N G S P G S Y A V N T F N O N L S D I A F B E S G A K Y T S D E M G T I O F K P D E V L I O N G F T S O A R N E			1015
Dd	1012	S T F S D P R S M I A T O A N - - - -	C Y T Y - - - -	I G N A L	1038
OY	1019	V T N O S F I N - - - S L V D F T P A N A G T N Y R V Y D P D G N L T N O N L P L K V O I Q I Y I D D K R Y A K L K N			1074
Dd	1039	V S N I G S I D T P A S V R F T G C N D S G A L Q - - - -	C N I Y S O N - - - -	I D F G T Y N L T I I N	1083
OY	1076	N L V - - - - T E S Y N N - - - - -	F A L P S W V V P A I C S T L G I		1104
Dd	1084	S N V I I G G T T A I N E I D L L T N N L I F A N G S T W G D N S I S I T L A V			1127
RESULT 5					
US-09-206-942-67					
; Sequence 67, Application US/09206942					
; Patent No. 643269					
; GENERAL INFORMATION:					
; APPLICANT: Loosmore, Sheena M.					
; APPLICANT: yang, Yan-Ping					
; APPLICANT: Klein, Michel H.					
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High					
; FILE REFERENCE: 1038-861 mts:lp					
; CURRENT APPLICATION NUMBER: US/09/206,942					
; EARLIER FILING DATE: 1998-12-08					
; EARLIER APPLICATION NUMBER: 09/167,568					
; EARLIER FILING DATE: 1998-10-07					
; NUMBER OF SEQ ID NOS: 95					
; SOFTWARE: patentln Ver. 2.1					
; SEQ ID NO 67					
LENGTH: 1536					

```

Db      911 TINTFVAGLFDKGN-----ISIAKGARFDDIDSKNISTTSSSTRTI 960
QY      887 FPGNOLWYLETNEN-----NKSSTYTLRL-----911
Db      961 ISGN-----ITRKNBDLNTNEGSDTEMOIGDVSQKEGNTLSSDKINIKQITIKAG 1014
QY      912 ADSNDASSSFPTLIDVNEIGVILPLIDNSF-----YVNAAGVNALESSNPGSGSY 967
Db      1015 VDEGSDSDATNNANLITIKELKLTQDLNISGFNKAETAKGSDLTIGNTN-SADGYN 1073
QY      968 TAVTFNQNLSIDIAFEGSGAKYSDPWGTTQKFPDEYLIONGFTSQARNFYNOSPLNS 1027
Db      1074 AKKVFENQ-VKDSKISADGKHVT-----LHS 1098
QY      1028 LVDETPANAGTNRVVVDDGNTLTONLPLKYOIYLDGKYDAK--LKNNNLVTFESYN 1085
Db      1099 KVTSGSSNNTE-----DSDNNAGLFI-----DAKNVTYNNNITSHKAVS 1139
QY      1086 FGLPSPWYV-----TAIGSTIGLAIMIILGLAIGIPLRAQRKLODKGFTTKKVDTLTA 1142
Db      1140 ISATSEITTKTGTNNATTGAVNEITAGTGSILG-----GIESSSGSV-TLTA 1186
QY      1143 -----AVGSYKKIITOTAN-----VKKRPAALGAGKSGD 1172
Db      1187 TEGALVNSNSGNTVYTNANSALTLTAGSTIKGTESVTTSSQSGD 1232

```

```

RESULT 7
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
;
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match      3.0%; Score 190; DA 1; Length 2710;
Best Local Similarity 19.3%; Pred. No. 0.00029;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;

QY      42 VNQARTLDANSVRLAGLQNGSLFNVLVDVDFITPAANGTIKIDSTFKPL-----94
Db      1071 INMSLSIAATVASIVIGAEVTFILPLIAGISGIPSLVNNELIHDKATSVNVYFHL 1130
QY      95 -----YGLSDDCGKVKVQIVSDYTTSSNRREDQKQTRAYVALVNDENVHLKRI 149
Db      1131 ESKRYGPKLTEDD-----KILVPIIDLVLSEIDF-----NNN 1162
QY      150 SNRIGNRN-----NNSKFVIGGVDPNPAHVIREFTDGTFENOTGEIVNDFILDAP 203
Db      1163 SIKLGTNCILAMEGSGSHYVTGNIDH-----FFSSPSISSHPLSLISYSA-- 1207
QY      204 LPKDLHPDWYNLYIQRKI--LPNDVNTAVVWPVGRV-----SGTNADGMPGCGNG 253
Db      1208 -----IGIETENLDFSKIKMLPN-APSRVFMWETGAVGLRSLNDGTRLLDSIR 1262
QY      254 Q-----ITNDPIAO-----TKTTDQNSTNSGAMP 283
Db      1263 KFYWRVYAFPDVATTLKPYVEDTNKIKIKDKOTRNFIMPTITTNELRNLSY---SFDG 1319
QY      284 ANNRYSQNLNVKRIKTSFQDLDERINTNSRIGNRNNNSKFVIGGVDPNPAHVIREFTDGT 343
Db      1320 AGGTYSILL-----SSYISTININLSKQDLM-----IFNIDNEVEIENGTL 1363
QY      344 KFNFTNOTGEIVNDFILDAPILPKDLHPDWYNLYIQRKILLPNDVNTAVVWPVGRVSGT 403
Db      1364 K-----KGLIKD-----VLSKI---DINKNKL-----ITGN 1387
QY      404 NADGMPDGCNGGQITNDPIAQTKTTDNONPSTFN-----SGAMPANNRYSQNL- 454
Db      1388 QTTD-----FSGDINDKDRYIFLTCELDQXISLIIEINLVAKSYLLSDDKNYLLSNLS 1442
QY      455 NVKRIKTSFQDLDEKFEYYPWMTGSEENKNTRLATGSLPSENERVWILDIGTPQVTLKED 514
Db      1443 NTEIKINT-LGDSKNIAVYTTDESNNKIGALSKITSQKIIHY-----KRD 1488
QY      515 SVNVFSRLYNSV---NSLSFIGDSIYIFCTSELPSL---WYSFPTRLS-----DITAL 563
Db      1489 SKNIL-EFYNDSTLEFNSKDFIAEDIVFMKDDINTTGYVYDNNNDKSIDESISLVS 1547
QY      564 NOVKTDDIEASSTDNGTTNGTTTATSSGSGAGCGNTNTSQVSNPLTYNRYBSFGI 623
Db      1548 NOVXV-----NGLYLNEYSYSSYLDVFKNSDGHNTSNFNNFLD-NISFWKLGF 1597
QY      624 DSKPTSAK-----IDETNADPNVIEARIYAEVRLGIONIPITNAGFNRTNIGVGFT 679
Db      1598 ENINFEVDKFTLVGKNLGL-----YEFICDNKNKNDI-YFEWMTSSSKRTIFS 1647
QY      680 STGSRVVLRAASYNGDORPTG-----NFQPLVYFGVLGIGQOQRTTGTF-----721
Db      1648 GNGRNVVVEPIYND-----TGEDISTSLDFSEPLYGIDRYINKVLAPDLTYLSLININTN 1704

```

	REFERENCE/DOCKET NUMBER:	OPHD-01308
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE: (415) 705-8410	
	TELEFAX: (415) 397-8338	
	INFORMATION FOR SEQ ID NO: 6:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 2710 amino acids	
	-Type: amino acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: protein	
	US-08-405-496A-6	
	Query Match	3.0%; Score 190; DB 2; Length 2710;
	Best Local Similarity	19.3%; Pred. No. 0.00029;
	Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;	
QY	42 VNQARTDANSVRLAGIGNGSLFNTVLARDVDNFTFPAANGTIKIDSTFKL-----	94
DB	1071 INMSISIAIVASIVGIGAEVETIFLPIAGISAGISPLVNNELIHDKATSVYVNFNHLIS	1130
QY	95 ----YG-ILDLSDCGGKRVKQIVSDYTSNRRFDQROTGRAVYALVLYDEANVHLKRITNN	149
DB	1131 ESKKYGPKLTEDD-----KILVPIDDLYISELDF-----NNN	1162
QY	150 SNRIGNRN-----NNSKFVIGVDNPAHYIRFTDGTGKFNFINQNGEIVNDFILDPI	203
DB	1163 SLKLGTCNIIAMEGSGHVTGNIDH-----FPSSPSISSHPSLSIYSAA--	1207
QY	204 LKRDILHPWYNLYIORKI--LPNDVNFVPMVPGRY-----SGTNADGMFDCGNG	253
DB	1208 ----IGIETENIDFSKKIIMLPN-APSKVPMWETGAVGLRLENDGTRLDLSIRDLYPG	1262
QY	254 Q-----TNTDPIAO-----TKTTDONONSTENGSCAMP	283
DB	1263 KFYWFYAFEDYAIITLKPVEDYEDTIKIKLDKTRNFMPTITTNELRNKLSY---SFDG	1319
QY	284 ANNRYSQNLNVKHKIKTSPQDERINTNSRIGNNNNSKFVIGVDNPAHYIRFTDGT	343
DB	1320 AGGTYSLL-----SYPISTINILSKDML-----IFNIDNEVREISIEGTTI	1363
QY	344 KFNFTNOTOGELVNDFILDAPIIPKIDHPWNLYIORKILPNDVNFVPMVPGRYVSGT	403
DB	1364 K-----KGLIKID-----VLSKI--DINKKL-----IGN	1387
QY	404 NADGMFDCGNGOILTNPIAQTKTITDONONSTEN-----SGAMPGANRRYSOL-	454
DB	1388 QTID-----FSGDIKDKRATFELGCELDKISILEINLVAKSYSILSGDKNYILSLS	1442
QY	455 NVKHKIKTSPQDERKRYFPMWETSEENKNTIRLATGSLPSNRURYILDIPGROYLTLED	514
DB	1443 NTIEKINT-IGLDSKNIAWYTBESNNKIRGAISKISQSIHH-----RKD	1488
QY	515 SVNFESRLYINSV--NSLSPIGDSIYFEGTSELPL--WYSPFTRLS-----DIAL	563
DB	1489 SKNII-EEYNSTLEFNSKDFIEDLVFEMKDDIINTGKYVDNNTDKSIDFSLISLYK	1544
QY	564 NOVKTIDIEASTONGTNTGTTTATTSAGTSAGAGTNTNSQTSVNPNTINTYRSFGI	623
DB	1548 NOVRY-----NGLYLNEVSYSILDEVKNSDGHNNNSPNNMLFLD-NISFWKLQER	1597
QY	624 DSKPTYSANK---IDETNMADPNVIEARITAEYKLGIONPIPTTNGNFRNTIGGVGFT	679
DB	1598 EMINIVNIDKRYFLVAGKTNIG-----YVEFICDNKKNIDI-YREKMTSSSKSTIIS	164
QY	680 STGSKRYVLASVNGQORPTG-----NFQPLVYFGYLGQOQRTGTF-----	721
DB	1648 GNGRNVVVEPIYNPD---TGEDISTLDSEVPELYGIDRYINKVLAPLDYLSLININTN	170
QY	722 ----WYGTGLNNSPYD-----LDSP-----RVGNETNFRRTSLTYRPMVGXYLDEEG	767
DB	1705 YYSNEVEYPIILVNPTEPRKKVKNINLDSSSEFYKWSSTGDSF-----ILVRYLEE--	1751
QY	768 AKSFSTNPIIRACQDTPERSIFGQSYSDNTYVEYIQSVLGFEDGIRNNLNGVAKASSPLNS	827


```

Db 1755 -----SNKKIL-----OKIRIGILSNTQSPNMSIDFDIK-KLSLGYIMSKFSP 1800
QY 828 NRPNP-----NGLEMAATTYLRSQ-----IGLARTSGL-----PNOQPFQTHOYIS-- 870
Db 1801 NSENELDRDLKFKIDNKTYYYDEDSKLVKGLININNSLFFYDPDIEFLVGMQRIK 1860
QY 871 -----VSPGDQSSIKNIRTIFFPGNQLWELFTNENKSSVYTLRLADSDNPASSPSP 925
Db 1861 KYEFDINTGALITSK-----LNGK-----HFYRNDG-----VMOLGVTKPGDFEYFAP 1907
QY 926 TSLIDVNEIGVILLPLDINSFYVNAAGNALFSSNPGSYAVVT-----FQNLSDLA 981
Db 1908 ANTQNNNIEGQAI-VYQSKFELTN--GKKYFDNNSKAVTGRIINEXKYFNP-NATA 1963
QY 982 FEG-----SGAKY-----TSDFMGTIO-----FKPEYLIONGTSQVARNFYTNQ-- 1022
Db 1964 AVGLQYIDNNKYFFNPDFAIISKQOTVNGSRYYEPTDPAIAFNGYKTIIDGKHFFYEDSDC 2023
QY 1023 -----SFLNSLVDFTPANAGNYKVVVDPGNLNOMLPLKVOIQYIDGK--YYDAKL.1073
Db 2024 VKIGYFSTSGNEFEFAPANTYNN-----NIEGQAIYQSKFELTNKGYTFD--- 2071
QY 1074 KNNNLVT-----FSYNNFGALPSW 1092
Db 2072 NNSKAVTGLQTIIDSKKYYENTMTAEATGM 2101

```

```

RESULT 9
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6230960
;
; GENERAL INFORMATION:
;
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIORITY APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: 0PHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-915-136-6
;
; Query Match
; Best Local Similarity 19.3%; Pred. No. 0.00029;
; Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;
;
;
; 42 VNQARTLDANSVRLAGLNGSLFNTVLRVDNDNFITAANGTIKIDSTFKPL----- 94
; 1071 IMMSLSIAATVASIVGAEVITFLPIAGISAGIPSLVNNELIHDKATSVVYENHLS 1130
; 95 -----VG-LDSDCCGGYKVKQIVSDYTSRRNRPDQGTBAYVALVNDENAVHLKRINTN 149
; 1131 ESKKIGPLKTEDD-----KLIVPIDLVISEIDF-----NNN 1162
; 150 SNRIGNRN-----NNSKFEVIGVDNPAHVIREFTDGFENFTNOTGELVNDFILDAPI 203
; 1163 SIKLGCNILLAMEGSGGHVTVGNIDH-----FFSSPSISSHIPSLISYSA-- 1207
; 204 LPKDLHPDYNLYIQKTI--LPNDVNTAVVPMVGVRY-----SGTNADGMEFCGNG 253
; 1208 ----IGIETENLDFSKIMMLPN-APSRVFMETGAVPGLRSLNDGTRILLDIRLYBG 1262
; 254 Q-----ITNTDPIAQ-----TKTTDNONPSTFGAMP 283
; 1263 KFYWRFAFDVAITLTKRYVEDTNKIKILDKTRNFIMPTTTNTRKLSI---SFDG 1319
; 284 ANNRYDSQNLVKHRIKTSFQDERINTNSRIGNRNNNSKFVIGVDNPAHVIREFTDGT 343
; 1320 AGCTYSILL-----SSYPISPINIMLSKDDLM-----IFNDINEVEISEINGRT 1363
; 344 KENFTNOTGELVNDFILDAPILPKDLHPDYNLYIQKTI--LPNDVNTAVVPMVGVRS 403
; 1364 K-----KGKLIKD-----VLSKI--DINKRKL-----IIGN 1387
; 404 NADGMEFCGNGQITNTDPIAQTKTTDNONPSTFN-----SGAMPGANRRYDSOL- 454
; 1388 QITD-----FSGDIDDKDKRIYFLTCELDDKISILIEINLYAKSYLLSGDKRYLLSNLS 1442
; 455 NVKHRIKTSFQDERKFEVPEMGSEENKNTIRLATGSLPSNERYWILLDIPGTPOVYLKED 514
; 1443 NTEKINT--LGDSKNIAVYTDENNNKFGAISKTSQSIHY-----KKD 1488
; 515 SVAVESRLYLSV---NSLSFGDSIYIGTSLPSL---WYISFPTRLS-----DITAL 563
; 1489 SKNII--EFYNDSTLFEKSKDIFEDINVEFKDDINITGKYVYDNTDTSIDPSISLYSK 1547
; 564 NOVKTDDIEASSTDNCGTTTNGTTTADTSGSGAGAGNTNTSQTYSNLTLYRSFGI 623
; 1548 NOVKV-----NGLYLNESYSSYSLDPEVKKSDGHHNTSNMNLFLD-NISFMKILGCF 1597
; 624 DSKPTSANK-----IDETNADPNVIEARIYAEVRLGIONEIPITNAGNFRNTIGGVGT 679
; 1598 ENINNFYIDKYFTLVKRTNG-----YVEICNNKNKIDI-YGEMKWTSSSKSTIFS 1647
; 680 STGSRYVLAASYNGDORPTG-----NFQPLLYVFGYLGQQRITGTF----- 721
; 1648 GNGRNVVVEPIYMPD---TGEDISTSLDFSEPLVGLIDRYINRVLLAPLDLYTSLININTN 1704

```

Query Match	3.0%:	Score 190;	DB 4;	Length 2110;
Best Local Similarity	19.3%:	Pred. No. 0.00029;		
Matches 238;	Conservative 155;	Mismatches 459;	Indels 378;	Gaps 61;
42	VNCARTLDANSVRLAGIUNGSLFNVNVADVDDNFITANGNIIKLSFTKPL-----	94		
Db	1071 INMLSIAAIVASIVGIGAEVTEFLPIGISAGISPLVNNELIHKDATSVVNFNLS	1130		
Qy	95 ----YG-LDSDDCGGKVKQIYSDYTSKNRPDQROTAYALLVNDENAHKRIKTN	149		
Db	1131 ESKRYGPKLEDD-----KILVPIDDLVISEIDF-----NNN	1162		
Qy	150 SNRIQNN-----NSKPYIGGVNDNAHIRTDOGTKEFNTOGELVDFILDPI	203		
Db	1163 SIKIGTCNIIAMEGSGHIVTGNIDH-----FFSSPSISHPISLSTYSA-	1207		
Qy	204 LPKDLHPWMIYIYORKI--LPNDVTAVVPMVGRV-----SGTAADGMDCGNG	253		
Db	1208 ----IGITENLDFSKIMLPLN-APRPFVWETGAPGLRSLENDGTRLDSIRDLYPG	1262		
Qy	254 Q-----ITNDPIAO-----TKTTDNONPSTFNSGAMFG	283		
Db	1263 KYWPFYFAYEDYAITTLKPYVEDINIKIKDKDORNFIMPTITNEIRKLASY---SFDG	1319		
Qy	284 ANNRYDSOLNKHRIKTSFOLDERINTNSRIGRRNNNSKFYIGGVNDNAHIRTDDGT	343		
Db	1320 AGGTSLLL-----SSYPISTINISKDDLM-----IFNDINEVRELISNGTL	1363		
Qy	344 KKNFTNODGELVNDPILDAPILPKDLHPWMIYIYORKIILPNDVTAVVPMVGRVSGT	403		
Db	1364 K-----KGLIKD-----VLSKI--DINKNKL-----IIGN	1387		
Qy	404 NADGMEFGCGNOIINTDPIAOIKTTTNDONPSTFN-----SGAMPANNRYSOL-	454		
Db	1388 QITD-----FSGDIDDKRYIFLTGELDDKISLIEINLVAKSYSLLSGDKNLLSLS	1442		
Qy	455 NVKHRIKTSFOLDEKFEVPEWMTGSEBKNITRLATGSLPSNERKWIIDIGTPQVTLKE	514		
Db	1443 NTLEKINT-LGDSKNIAVNYDESNNKYFGAISTOKSLTH-----KRD	1488		
Qy	515 SVAVFERKLYNSV---NSLFTGDSYIYIGTSELPSL---MYSPPTRLS-----DIAL	563		
Db	1489 SKNLL-EFYNDSTLERNKDFIADINVMKODINMTIGKYYVDNTEKSIDFISLVS	1547		
Qy	564 NOVKTDDIEASSTDNGTTNGTTTADTSSGSGTAGCTGNTNFTQSVNPLNTYRSFI	623		
Db	1548 NOVXV-----NGLYLMESSYSLDFVKNSDOHHNTSMFMFLD-NISPMKLGFE	1597		
Qy	624 DSKRTSANK---IDETNMADPVIIEARIYAELRIGIONEIPITNAGNIRNTIGGVGT	679		
Db	1598 ENINFEVIDKFTYLGRKTNLG-----YVEFICDNKNNDI-YFEGEKTKSSSKSTFS	1647		
Qy	680 STGSRVYLRAISYNGDQPTG-----NEOFELVPGFLYQGTQRTGT-----	721		
Db	1648 GNGRNVVVEPIYNDP---TGEDISTSLDSYELPIDRYIKVLLAPRLYSLINMTN	1704		
Qy	722 ----WGTYKLNNSEYDV-----LDSP---RVGETNQERKISLYTPWVGCLTEBG	767		
Db	1705 YYSNEYEILVLPNPFHKKVINILNDDSSFEYKWSSTEGSDF-----LLVRYLEE--	1755		

QY 768 ARSFNTPYIRAGDTPESRSIFQSGYSDNTVEYIOSVLGFDGIRNNLANVKAASPLNS 827
D 1755 -----SNKKIL-----OKIRIKGLISMTQSPFNKMSIDFKRIK-KLSIGYIMSNKSF 1800
QY 828 NRNP-----NGLEMTAATYLRQ-----IGLARTSGL-----PNOQPEGTHQVYS- 870
D 1801 NSENEIDRDLGKFIIDNKTYEYDEDSKLKVLGININSLEFDPLEFNLVTQMGTNGK 1860
QY 871 -----VSPGQFSSINIRTIIFPGNOLMYFLFTNENKSSVYLRDLADSSNPQASSFSF 925
D 1861 KYFEDINTGALYSK-----INGK-----HEFYNNDD-----VMQIGVEKPGDGFEXFAP 1907
QY 926 TSLIDVNEIGVILPLDINSFYVNAAGNVALLFSSNPGSPSYTAIVT-----FNQNLSDIA 981
D 1908 ANQNNNIEGOAL-VYOSKFLTLN--GKTYFNNNSKAVYGMILINNEKTYFNP-NATA 1963
QY 982 FEG-----SGAKY-----TSDFWGTIO-----FKPDEXLONGFTSQVARNFYTNQ- 1022
D 1964 AVGLQYIDNNKYFENPDTALISKQWOTVNGSRXYFDPDFAIAFNQYKTIQDKHFEYDSDC 2023
QY 1023 -----SFLNSLVDETPANAGTNRVYVDPDGNLTNOLPLKVOIQYLDGK--YYDAKL 1073
D 2024 VAKIGVSTJNSNGEYFAPANTYNN-----NIEGOALVYOSKFLTLNGKKXYED-- 2071
QY 1074 KNNNLVT-----FSYNNFGALPSW 1092
D 2072 NSKAVTGLQTIQIDSKKYFNTNTAEATGW 2101

RESULT 11
US-10-011-366-6
Sequence 6, Application us/10011366
Patent No. 6573003
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medien & Carroll
STREET: 220 Montgomery Street, Suite 2200
City: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/011,366
FILING DATE: 16-No. 6573003-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6
Query Match 3.0%; Score 190; DB 4; Length 2710;
Best Local Similarity 19.3%; Pred. No. 0.00029;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;
QY 42 VNQRITDANSVRLAGIGQSGLEFNYLVKRYVDNFTIANGTIIKIDSFKPL----- 94
D 1071 INMSLSIAATVAVSIGAEVTLFLPLIAGISAGISLVNNEILHDKATSVNYFNHLS 1130
QY 95 -----YGLDSDCGGKVKQIVSDYTSRNRFPDQROTAYVALVNDANVHLKRIINTN 149
D 1131 ESKKYGELKTEDD-----KILVPIIDLAYSIDF-----NNN 1162
QY 150 SNRIGNRN-----NNSKEVIGVDNPAHAVIRFTDGTCKFNFTNQTGELVNDPILDAPI 203
D 1163 SIRLGTGCIILAMEGSGHTVTGNIDH-----FFSSPSISHSIPSLSYSA- 1207
QY 204 LPKDLHPDMWNLVYQRI--LPNDVNTAVVPPWPGVRY-----SGTNADGCFDGCNG 253
D 1208 -----IGIETENLDPSKKIMLPN-APSRVEMETGAVGLRSLSDSTRLDLSRDLYPG 1262
QY 254 Q-----ITNTDPIAQ-----TKTTDQNPSTFNSGAMP 283
D 1263 KFYWRFAFFDYAITTLAKPYEDNTIKIKDKDQTRNIMFTITNELRNLSY---SFDG 1319
QY 284 ANNRYDSQLNVAKRIRIKTSFQLEDERINTNSRIGNRNNNSKFVIGVDNPAHAVIRFTDGT 343
D 1320 AGGYSLLL-----SSYPISYINILSKDILW-----IFNIDNEVEREISENGTI 1363
QY 344 KFNFTNQTGELVNDPILDAPILPKDLHPWNLVYQRI--LPNDVNTAVVPPWPGVRYSGT 403
D 1364 K-----KGLKLD-----VLSKI--DINKNKL-----ITGN 1387
QY 404 NADGMEDCNGGQITNTDPIAQRTKTTDQNPSTFN-----SGAMPANNRYDSQL- 454
D 1388 QTID-----FSGDIDNDRYIFLTCEDDKISLIEINLVAKSLSLLSDGKNLISNLS 1442
QY 455 NVKRIKTSFQLEKFEYVPEMTGSEENKNTIRLATGSLPSNERVWILDIGTPQVTLKED 514
D 1443 NTIEKINT-LGLDSKNIAVNYTDESNNKRYGALSISKTSOKSIINH-----KRD 1488
QY 515 SVNVFSRLYLNSV---NSLSFIDSLYIEGTSELPRL---WYVSFPTRLS-----DTAL 563
D 1489 SKNIL-EFYNDSTLEFNSKDFIAEDIVFEMKDDINTTGGYYVYDNNDKSIDPSISLVSK 1547
QY 564 NOYKTDIEASSTDNGTTNGTTTADTSSGSGAGCTNTNTSOTYSNPLNTYRSFGI 623
D 1548 NOYKVV-----NGLYLINESYSSLYDVKNSDGHNTSNFNLFLD-NISWKLFGF 1597
QY 624 DSKPTSANR-----IDETNMAPNVYIEARVIAEYRLGIQNEIPIITNACNFIKNTIGVGFT 679
D 1598 ENINFIKDYFTLLGKTNLG-----YVEFICDNKNKNDI-YFCEWMTSSSKSYIFS 1647
QY 680 STGSRVVLASYNDDQRPFG-----NFQPLLYVFGVLQGTPTGTF----- 721
D 1648 GNGRNVVVEPIYND--TGEDISTSLDFSEYEPYGGIDRYINKVLAPDLTYTSLININTN 1704
QY 722 -----WYGTYKLLNNSPYDV-----LDSP-----RVGTEYQFRSTSLTYVPMGGLTEEG 767
D 1705 YSNVEYPELLIVLNPNTFHKKVNININDSSFEYKWSYEGSGDF-----ILVRYLEE-- 1754
QY 768 ARSFNTPYIRAGDTPESRSIFQSGYSDNTVEYIOSVLGFDGIRNNLANVKAASPLNS 827
D 1755 -----SNKKIL-----OKIRIKGLISMTQSPFNKMSIDFKRIK-KLSIGYIMSNKSF 1800

Query Match	3.0%;	Score 189.5;	DB 3;	Length 2353;	
Best Local Similarity	18.6%;	Pred. No. 0.00025;			
Matches	256;	Conservative 177;	Mismatches 479;	Indels 463;	Gaps 67;

QY	30	KQSDKNDNFOLVQARITLDANSYRLAGLQNSGLENTVLRLDYVDNEFTIARNGTIIKLIDS	89
Db	1141	KQSEK--DFTISLDITLTGLTSTLGGTANGNRDFTGVI-NKDGTLITLANGAAGATDA	1196
QY	90	FTKPLPYGLDLSDDCGGYK---VQIYSDTITSNRFDQOTRAYVALLVN-----	136
Db	1197	SMGNTISTYTKDGISAGNKEITNVKSLAKTYDONTADEYODKEFHAAVKNANNEVEYVG	1256
QY	137	-----DEANVHLKRINTNSNRI-----	153
Db	1257	NGATVSAKTDNNGKHTVTIIVAEAKVYDGLERODDGKIKLKVMDTDGNNLLTVDATKAS	1316
QY	154	-----GNRRNNSKFPYIGGVNDPAHYIRFTDDGTFKNFNPNOGVELYN	195
Db	1317	VAKGEFNATVTATTAAGGNANERKKVYVVG-----SNGATATETDKKRVATVG	1365
QY	196	DF---TLDAPILPKDLHPD-----WYNLTIOKK-----	220
Db	1366	DVAKALINDATAFYKVENDDSATITDSPYDDGANDALAKAGDTLLTKAGKNLKYKRDGNKIT	1425

QY	221	-ILPN--VNFVEMPV-----GRSGTNADDG---PDCNGQITN--IDPLAQFT	266
Db	1426	FALANDSVKSAVSDKSLGTNGNKKVNTDITDGLNFARDKSTGDDANHLGLASTLT	1485
QY	267	TTDNONSPTEPNSGAGANNRRYDSOLNKHRIKTSFOLDERINTNSNRGNRNNSK---	323
Db	1486	DTLNSGATTLNAGCITDNEKKRAASVDVLNMGW-----NRYGVKPAASANNQVENID	1539
QY	324	-----FVIGVDNPAHVITRFTDDGTRKPFNTQNGEIVNDFILDAILPKLHPDM	374
Db	1540	FVATYDVDFVSDGDKDTSTVYESKDNKGRKEVKGAKTSYIKR-----	1583
QY	375	YN--LYIQRIKLPNDVTAFAVPMVGRVSGTNADDGMFCNGQITNTDPLA-----Q	425
Db	1584	HNGKLFGKELKADANNNGVTVTEBDG-----DENGGLVATAKAVDAVKNAGMR	1632
QY	426	TKTYYDN-QNS-----TF--NSGAGANNRRYDSOLNKHRIKTS--FOLDEKT	470
Db	1633	VKTGANGQNDDEFAVVASGTNTVFAFGNGTTFAYVKANDGSITVYVKNVADGKLIDGDK	1692
QY	471	VYPEWTGSEENKNIITRLATGSLS-----NRRYIIDI--PGTPOV	509
Db	1693	IYADT-----VLTVAADGKVTAFAPNNGDKREVDASGLADLINKLSTAFARAGEGGEV	1745
QY	510	TLKESVAVFRLINLSVNSLSF--IGSITFEGTSELPLMWYSFPRLSDLTALNOYKT	568
Db	1746	-----DPNASSAQ--EYKAGDKVTFEAGNLMKI--KOSGDEFTYSILKELKDLTS-----	1791
QY	569	DDIEASSDNGT-----TNGTITTAADTSGSGTAGGNTNTSQ-----	608
Db	1792	--VEFDANGGTGSESESTKITPKDGLTTPANGAGAGANTANTISVTKDISAGNKAVTNV	1849
QY	609	-----TVSNPL-----NTYRSF-----GIDSKPTSANKIDET-----	636
Db	1850	VSLGKFFGCHLANGTVADEFKHYDAXKDLTILDEKGDANNPTADMTAATVGDGLK	1909
QY	637	NW-----ADPNVTEARITAEVYLGIG--NEPIPTNAGNFTRNIGGVFTSGSRVY	686
Db	1910	GWVTSADKTTGEBN-----QETMAOYRANNEVKFKSG-----NGINSGKTLNCTRYI	1957
QY	687	LRAVYNDQRPNGFNPOPLVYFGLYGOQ--TRGTGFWYGYKLLNNSPVDVDSRVCY	744
Db	1958	TTELAKGEVYKSNEFT--VKNADGSEFNLYKKGDMY-----SKEDIDPATSKPMTG	2007
QY	745	ETNOFR-----RTSLTFPMVG--GYLT--EGGARSESNTPYTRAOGDTPESRSIF	790
Db	2008	KTEKYKAEKNGKVVASANGKTEVTLTKNGSGYVTVQGVADATAKGFEGLADAAEAKAF	2067
QY	791	QSGYSDN--TYEYIOSYLGFDCIR--NNLNNGVKKSSFLNSRPNGLNEMIAATYLR	846
Db	2068	AESAKDQLSADKRAETVNAHDKVRPANGTNTVSAIYEST--DANGDKV--TTFEVT	2122
QY	847	QIGLARSGPLNOOPFGTT--HOVISVSPGDOFSSIKNIRTIIFPGNOLWFLPTNENK	903
Db	2123	DV-----ELPQIOTYNTDANGKRIYKAKDGK-----WELNADGTAS	2159
QY	904	SSVYTLRLADSSNPDASSSFPSISLIDVNEIGVILPLDNSEYTVNAAGNVALEFSSNBS	963
Db	2160	NKEVTL-----GNVDANKK--KVAVYTLNGA--DKWYTT--NADG-----	2193
QY	964	PGSYAVNTVFNQNLSDIAFEKSGAKYTSDFWGTIOKRPDE-----YLONFTSOVARN	1017
Db	2194	-----AADTKG-----EVSNDKSVSTDEKHVYRLDPNNSGNGKGVADIVANAEIS-	2239
QY	1018	FVTNOSFLNSLDFTPANAGTNYRVVVDPGNLTNQLPLKVOIOYLDGRYYDAKLKNN	1077
Db	2240	-ATSDAINGSOLYVAVAGVTN-----LAGGVNLTERRKVRKVQ--KRAD	2280
QY	1078	LVTFEYVNNGALPSWVPTAIGSTGLIATMIL--GLAIGTFLAOKRLDDKG	1129
Db	2281	AGTASMLASQLPQATMP-----GKSVUATAGSSYOGONGALATGV-----SRISDNG	2237

RESULT 13
US-08-913-942-4
Sequence 4, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RT/RMS/DAY
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 27299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

Query Match 3.0%; Score 189.5; DB 3; Length 2353;
Best Local Similarity 18.6%; Pred. No. 0.00025;
Matches 256; Conservative 177; Mismatches 479; Indels 463; Gaps 67;

QY 30 KQSDSNDNTOLVNAKRTLDANSVRLAGLGNGSLFNTVLRVDNFTFAANGTIKIDS 89
DB 1141 KQSEK--DFTYSIQDTLT--GLTSITLGGTANGRNNDGTVI--NKDGLITLILANGAAGTDA 1196
QY 90 FTRPLXGLDSDCCGKY--VKQIVSDYTSRNRFPDOROTRAYALIVN----- 136
DB 1197 SNGNTISVTKDGISAGNKEITVKSALKTKYTQNTADETQDEFFAAVKNANEVFTVG 1256
QY 137 -----DEANVHLKRIINTSNRI----- 153
DB 1257 NGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTQKIKLKVDNTDGNLLFTVDATKGAS 1316
QY 154 -----GNRNNNSKFVYIGVDNPAHVIRPTDDGKFKFNTQGEIYN 195
DB 1317 VAKGEFNAVTTDATTAQGTNANERGVVYVG-----SNGATATETDKKVAATVG 1365
QY 196 DF---ILDAPILPKDLHPD-----WYNLYIQRK----- 220
DB 1366 DVAKAINDADATFYKVENDDSATIDDSPTDGANDALAKAGDTLTLLKRGKMLKVKRDSKNTT 1425
QY 221 -ILPND--VNTAVVMPV-----GRVSGTNADDGN--FDGCGQITN--TDPIAOTKT 266

DB 1426 FALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLNFAKDSKTGGDANIHLNGLASTLT 1485
QY 267 TTDNQNPSTFNSGAMGAMNRRYDSQNLVHRIKTSQDDEIRNTNSNRNGNRRNNK----- 323
DB 1486 DTLLNSGATTNLGGNGITDNEKKRAASVDVILNAGV-----NVRGVKASANOVENTID 1539
QY 324 -----FVIGVDNPAHVIRPTDDGKFKFNTQGEIYNDFILDAPILPKDLHPDM 374
DB 1540 FVATIDTVFVSGDKDITTSVYESKDNKRTVEKIGAKTSYIK----- 1583
QY 375 YN--LYIQKKILPNDVNTAVVMPVGVSGTNADGMDGCGNGQITNTDPIA-----Q 425
DB 1584 HNGKLETFGKELDANNNGVTJETDQK-----DEGNGLVTAKAVIDAIVNKAQR 1632
QY 426 TKTTIDN--QNPS-----TF--NSGAMGAMNRRYDSQNLVHRIKTS--FQDEKF 470
DB 1633 VRTTGANGONDDEFATVASTVNTFADNGCTTAETVRKANGDSITVKYNVAVDGLKLDGDK 1692
QY 471 VYPEWTSBENKNITRLATGSLPS-----NERYWILDI--PGTPQV 509
DB 1693 IYADTI-----VLTVADGKYTAPNNGDKKFEVDASGLDALNKLSTWATAGKEGTGEV 1745
QY 510 TKEDSVNFSRLYNSVNSLSE--IGDSIYIFGTSLSPLWYSEFPTRLSDLTALNQVKT 568
DB 1746 ---DPANGAQ--EVKAGDKVTFKAGDNLKI---KQSGKDFYSLKKEKLDLTS----- 1791
QY 569 DDIASSTNGT-----TTNGTTTATSSSTGAGTGNNTTSQ----- 608
DB 1792 --VEFKDANGGSGSESTKITKDGTLTPNAGAGAMANTANTISVKKDISAGNKAVTNV 1849
QY 609 -----TVSNPTL-----NTRSF-----GIDSKPTANKIDET----- 636
DB 1850 VSGLKKFGGHTLANQTVADFEKHYNAVKDLNLNDEKADANNPTIADNTATVCDRL 1909
QY 637 NW-----ADPNVIEARIYAEVRLGIO--NEIPTNAGNFIRNTIGGFTSTGSRVY 686
DB 1910 GHWISADKTTGEBN-----QEVNAQVNAHEVKKSG-----NCINVSGLTLNCTRYI 1957
QY 687 LRASVNGDQRPNGNFQPLFYVFGYGLQO--TRTGFWYGTLYLLNSPYDLSRVGT 744
DB 1958 TRELAGEVKSNEFT-----VKNADSENLVKVGMVY-----SKEDIDPATSKPMVG 2007
QY 745 ETNOFR-----RTSLTPYPMG--GYLT--EKGARSFSTPYRAOGDPESRSIF 790
DB 2008 KTEKIVENGKVVASANGSKTEVTYLTNKGSGYVYGNVADAIKSGELGLADAEEKAF 2067
QY 791 QSGYSDN--TYEYIOSVLGFDGIR--NNLNVGVKASSFLNSRPNNGLEMIATYTLRS 846
DB 2068 ABSAKDKQSKDKAEFTVNAHDKVRFANGLINTKVSATVBSST--DANGDKV--TTTFVKT 2122
QY 847 QIGLANTSGLPNQPPGCT--HQVIVSPBGQFSSIKNIRITFPENQMLYFLTENNKK 903
DB 2123 DV-----ELPLOIYINTDANGNKIVKKADGK-----WYELNAGGTAS 2159
QY 904 SSVYTLRLADSSNPDASSFPSLTLDVNEIGVILLDLNFSYTVNAAGVALFSSNPSS 963
DB 2160 NKEVTL-----GAVDANGK--KYVKVLENKA-----DKWYTT--NADG----- 2193
QY 964 PGSYTAVNTFNONLSDIAFEGSGAKYTSDFWGTIOQKPE-----YLIQNGFTSOVARN 1017
DB 2194 ---AADKTKG-----EVSNDKYSTBEKHVRLDPNNQSGKGVVIDNVANGCIS-- 2239
QY 1018 FVTNOSFLSYDFTPANAGTNRVRYVDDGNLTNQNLLPKVQIOYLDKAYDAKRLKNN 1077
DB 2240 -ATSIDALNGSOLYAAKGVTN-----LAGOVNNLEGVKNVGK--KRAD 2280
QY 1078 LVTFSYNNGCALPSPWVPTAIGSTGLILAIMIL--GAIGIPLRAQKRLDPKG 1129
DB 2281 AGTASLASASQLPQATMP--GKSMVAIAGSSVQGGNGLAIGV-----SRISONG 2327

RESULT 14

```

US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669,974
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match          3.0%; Score 189.5; DB 4; Length 2353;
Best Local Similarity 18.6%; Pred. No. 0.00025;
Matches 256; Conservative 177; Mismatches 479; Indels 463; Gaps 67;

OY 30 KOSKSDNTOLVNQAATITDANSVRLAGGONGSLPNTVLRVDVDFITANGTIKIDS 89
DB 1141 KOSER--DFYSLDDPILT-GLTSTTIGGTANGRNDGTIY-KNDGLITILANGAAAGTDA 1196
OY 90 FTKFLYGLDSDDCGGYK---YQIVSDYTSRNRDQRPATYALLVN----- 136
DB 1197 SNGMTISTYKDGISAGKKEITNVSAKTYKTONADETQDEFAAAYANANEVEVGK 1256
OY 137 -----DEANVHLKRIINTNSNRI----- 153
DB 1257 NGATVSAKTNNNGHHTYITIDVAEAKVGDGLEKPTDGKIKAKVNTGNNLLIYDARUKAS 1316
OY 154 -----GNRNNSKFVIGVDNPAHVIRFTDDGRTKFNTOGELYN 195
DB 1317 VAKGEFNAVTTDATAGCTANENERGKYVVG-----SNGATATETDKKKVATVG 1365
OY 196 DF--TLDAPILPRDLRD-----WYNLYIORK----- 220
DB 1366 DVAKAINDAATFEKENDDSATIDDSPTDGDANDALAGDTLITAKAKNLKVRDKNIT 1425
OY 221 -LLPND--VNTAVVPMV-----GRVSGTNADDGM--FDGNGQITN--TDFIAOTKT 266
DB 1426 PALANDLSYKSAVTSKISLGTNGKKNVITSDTKGLNFAKDSKGDADANILHNGIASLTIT 1485
OY 267 TTDONOPSTFNSGAMPGANNNRYDSOLNKHRIKITSFOLDEILNINSNRIGNRNNSK--- 323
DB 1486 DFLNSGATTNLGGGIGITDNEKKRAASVAKVLAAGW-----NVRGKVPASANNQVENID 1539
OY 324 -----FVIGVDNPAHVIRFTDDGRTKFNTOGELYNDFILDAPIPRDLRPDW 374
DB 1540 FVATYDTVDFVSGDKDITTSVYESKDNKRTVEKIGATSVYIKD----- 1583
OY 375 YN--LYIQRKILPNDVNTAVVPMVGRVSGTNADDGMDCGNGQITNDFIA-----Q 425
DB 1584 HNGKLETFGEKLDKANNNGVTJETDCK-----DEGGLVTAARAVIDAVNKAQWR 1632
OY 426 TKTTITN-QNPS-----TF--NSGAMPGANNNRYDSOLNKHRIKITS--FOLDDEFK 470
DB 1633 VKTTGANGDNDFAIYASGTNTVFADNGTITAEVIRKANDGSITVYKVKVADGKLDDGK 1692
OY 471 VPEWTGSEENKNITPLATGSLPS-----NERWILDI--PCTPOV 509
DB 1693 IVADDT-----VLIVADGKVAPANNQDGKRFVDSAGLADALNKLSTWATAGKEGTGEV 1745

```

```

OY 510 TLKEDSVNFSRLIYANSVNSLSF-IGDSIYIFGTSLELPSLVYSEFPTRLDPLANQVKT 568
DB 1746 ----DPANSAGQ-EVKAGDKVFEKAGDNLKI---KOSKDTFTYLSKKELKDLTS----- 1791
OY 569 DDEASTDNGT-----TTNGTTTADTSSGTGACTGNTTNSQ----- 608
DB 1792 --VEFKDANGTGSSESTKTKDKGLTITPANGAGAGANTANTISTYKGISAGKAVTNV 1849
OY 609 -----TVSNPTL-----NTRSP-----GISKPTANKIDET----- 636
DB 1850 VSGLKRFGDGHTLANGTVADEFKHDNAKDLTINDEKAGADNPVANAATVAGDLRGL 1909
OY 637 NW-----ADPNVTEARITAEYRLGIG--NEIDITNAGNFIIRITIGCVFTSGSRVY 686
DB 1910 GWYISADKTTGEPN-----OEYNAOVNANEVEFKSG-----NGIVSKTTLNGTIVY 1957
OY 687 LRASYNGDORPTGNEPFLXYCYLGYOQ--TRTGFWGTGYKLNNSPYDLDSPRYGT 744
DB 1958 TPELAGGEVVKSMET-----VKNADGSETNLYKVGDMY-----SKEDIDPATSKPMTG 2007
OY 745 ETNOER-----RTSLYYPVWG-GYLT-EGARSFSNTPYIRAQGTPEPSRIF 790
DB 2008 KTEKYKYENGVYVANGSKTEVTLTKKGSYVYGNQVADAIKSGEELGLADAAAEKAF 2067
OY 791 QSGYSDN--TYEYIQSYLCPDGIR--NNLNVGKASSFLNSRPNNGLEMIATTYLRS 846
DB 2068 AESAKDKOLSKDKAEYVNAHDKVFANGLTKVSAATVEST--DANGKV--TTTFVKT 2122
OY 847 QIGLARTSGLPNOOPFGTT--HOVISVSPGDPSSIKNIRITFPGNOLMYLFTINEKNK 903
DB 2123 DV-----ELPDIQIYNTDANGKNKIYKADGK-----WYELNDAGTAS 2159
OY 904 SSVYTLRLADSSNDPSSSPTSLLDVNETIGYLLPLDNEFYTVNAGVALLFSNPGS 963
DB 2160 NKEVTL-----GNVDANGK--KVRKYTNGA-----DKWYTT-NADG----- 2193
OY 964 PGSTAVNTFENQNTDLIAFEGSGAKYTSDFWGTIOFKPDE-----YLIONGTTSQYAKN 1017
DB 2194 ----AADKTKG-----EVSNDKYSTDEKHVYRLDPNNSQNGKGVVDINVANGEIS-- 2239
OY 1018 FVTNGSFNLVDEFTPANAGTNRVYVDPGONLNLNQLPLKVOLOYLDGKYDAKLKNNN 1077
DB 2240 -ATSTDAINGQLVAVAGVTN-----LAGOVNNEGKVNKVVG--KRAD 2280
OY 1078 LVTFESYNNFGALPSPWVPTAIGSTGLIAMIIL--GLAIGILPRAQRLODKG 1129
DB 2281 AGTASALASQLPQATMP-----GKSWVALAGSSYOGNGLAIGV-----SRISDNG 2327

RESULT 15
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-16
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match          3.0%; Score 189.5; DB 4; Length 2736;
Best Local Similarity 20.8%; Pred. No. 0.00032;

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 11:46:12 : Search time 800 Seconds
(without alignments)
10343.567 Million cell updates/sec

Title: US-09-901-572A-2

Perfect score: 3189
Sequence: 1 atgaatatctctaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published_Applications_NA.*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PC7_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PC7US_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3189	100.0	3189	US-09-901-572A-2	Sequence 2, Appli
2	3189	100.0	3189	US-10-131-591A-2	Sequence 2, Appli
3	3143.4	95.6	8354	US-10-125-818-1	Sequence 1, Appli
4	3057.8	95.9	3129	US-10-131-591A-79	Sequence 79, Appli
5	124.6	3.9	580073	US-10-205-220-1	Sequence 1, Appli
6	62.6	2.0	4985	US-10-056-405-10	Sequence 10, Appli
7	62.6	2.0	4985	US-10-094-240-10	Sequence 10, Appli
8	60.8	1.9	8771	US-10-311-455-1798	Sequence 1798, Ap
9	60.2	1.9	3489	US-10-294-804-1	Sequence 1, Appli
10	58.4	1.8	640681	US-09-790-988-1	Sequence 1, Appli
11	53.6	1.7	3673778	US-10-312-841-1	Sequence 1, Appli
12	52.4	1.6	3057	US-10-349-680-148	Sequence 148, App
13	51.6	1.6	2451	US-10-091-007-55	Sequence 35, Appli
14	50.8	1.6	2017	US-10-155-533-3	Sequence 3, Appli
15	50.8	1.6	5314	US-10-155-533-1	Sequence 1, Appli
16	50.2	1.6	2010	US-09-878-756-3	Sequence 3, Appli

c 17	49.4	1.5	1887	12	US-10-349-680-133	Sequence 133, App
c 18	47.8	1.5	3931	14	US-10-006-780-1	Sequence 1, Appli
c 19	47.6	1.5	1220	12	US-10-169-710-46	Sequence 46, Appli
c 20	47	1.5	575	11	US-09-918-995-6127	Sequence 6127, Ap
c 21	47	1.5	712	9	US-09-822-849A-186	Sequence 186, App
c 22	47	1.5	722	9	US-09-923-304-1	Sequence 1, Appli
c 23	47	1.5	722	10	US-09-880-107-1764	Sequence 1764, Ap
c 24	47	1.5	755	13	US-10-027-632-128109	Sequence 128109, App
c 25	47	1.5	888	9	US-09-925-302-116	Sequence 116, App
c 26	47	1.5	2405	11	US-09-296-523B-70	Sequence 70, Appli
c 27	46.6	1.5	6351	12	US-10-311-455-1717	Sequence 1717, Ap
c 28	46.6	1.5	6351	12	US-10-240-485-139	Sequence 139, App
c 29	46.6	1.5	8634	12	US-10-311-455-1029	Sequence 1029, Ap
c 30	46.6	1.5	3673778	12	US-10-312-841-1	Sequence 2, Appli
c 31	46.4	1.5	11155	12	US-10-311-455-578	Sequence 578, App
c 32	46.2	1.4	5641	12	US-10-311-455-1369	Sequence 1369, Ap
c 33	46	1.4	860	9	US-09-770-445-596	Sequence 596, App
c 34	46	1.4	5304	12	US-10-240-453-229	Sequence 229, App
c 35	45.8	1.4	5285	12	US-10-311-455-1753	Sequence 1753, Ap
c 36	45.8	1.4	5285	12	US-10-240-453-177	Sequence 177, App
c 37	45.8	1.4	5285	14	US-10-239-676-157	Sequence 157, App
c 38	45.8	1.4	6171	12	US-10-311-455-761	Sequence 761, App
c 39	45.8	1.4	6464	12	US-10-311-455-488	Sequence 488, App
c 40	45.8	1.4	6641	12	US-10-311-455-288	Sequence 288, App
c 41	45.8	1.4	6641	12	US-10-240-452-36	Sequence 36, Appli
c 42	45.8	1.4	15732	12	US-10-240-453-108	Sequence 108, App
c 43	45.8	1.4	15732	14	US-10-239-676-96	Sequence 96, Appli
c 44	45.6	1.4	6151	12	US-10-311-455-1584	Sequence 1584, Ap
c 45	45.6	1.4	6151	12	US-10-240-453-150	Sequence 150, App

ALIGNMENTS

```
RESULT 1
US-09-901-572A-2
: Sequence 2, Application US/09901572A
: Publication No. US20030165534A1
: GENERAL INFORMATION:
: APPLICANT: Nippon Zeon Co., Ltd.,
: TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
: FILE REFERENCE: J209
: CURRENT APPLICATION NUMBER: US/09/901,572A
: CURRENT FILING DATE: 2003-03-11
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 3189
: TYPE: DNA
: ORGANISM: Mycoplasma gallisepticum
: FEATURE:
: OTHER INFORMATION: mgc3 gene.
US-09-901-572A-2
```

Query Match	100.0%	Score 3189;	DB 12;	Length 3189;
Best Local Similarity	100.0%	Pred No. 0;		
Matches 3189;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGAATATTTCTTAACCAAACTTAACATTTATACATTTAGTGGTGGATTAGCTGATTTGGA	60	
DB	1	ATGAATATTTCTTAACCAAACTTAACATTTATACATTTAGTGGTGGATTAGCTGATTTGGA	60	
QY	61	GCTCTTGGTTCGCAAGCTTTGGCTTTAAGCAATCGATTAAGAGTAAACGATTAACACGCA	120	
DB	61	GCTCTTGGTTCGCAAGCTTTGGCTTTAAGCAATCGATTAAGAGTAAACGATTAACACGCA	120	
QY	121	TTAGTTAATCAAGCAAGCAAGCTAGATGCTAATTCGTTAGACTTCAGCTCTTGACAA	180	
DB	121	TTAGTTAATCAAGCAAGCAAGCTAGATGCTAATTCGTTAGACTTCAGCTCTTGACAA	180	
QY	181	AATGCTTCCTTCTTCAATACAGTTCTTAGAGATGTTGATGATTAACCTTTAATACAGCAGCT	240	
DB	181	AATGCTTCCTTCTTCAATACAGTTCTTAGAGATGTTGATGATTAACCTTTAATACAGCAGCT	240	

QY	241	AATGGAACAATTATCAAAATTAGACTGTTTACTAAACCATTAATATGTTTAGATCAAGT	300
Db	241	AATGGAACAATTATCAAAATTAGACTGTTTACTAAACCATTAATATGTTTAGATCAAGT	300
QY	301	GATGATTTGGTGGTGAATCAAAAGTAAACAAATAGTTTGAGATTACACAACTAGCAGAAAT	360
Db	301	GATGATTTGGTGGTGAATCAAAAGTAAACAAATAGTTTGAGATTACACAACTAGCAGAAAT	360
QY	361	AGATTTGATCAAGACAAACAGAGCATTTATGCTCTGTGGTTAAATGATGAAGCTAAC	420
Db	361	AGATTTGATCAAGACAAACAGAGCATTTATGCTCTGTGGTTAAATGATGAAGCTAAC	420
QY	421	GTTCAATTAAAGAAATTAATACTACTCAAAATAGAAATGTTGTAATGGAACACAAATCT	480
Db	421	GTTCAATTAAAGAAATTAATACTACTCAAAATAGAAATGTTGTAATGGAACACAAATCT	480
QY	481	AAGTTTGAATTTGGTGGTGGTGAATTCACAGCTCAGCTAATTAGATTACATGATGATGG	540
Db	481	AAGTTTGAATTTGGTGGTGGTGAATTCACAGCTCAGCTAATTAGATTACATGATGATGG	540
QY	541	ACTAAATTTAATTTTACAAACCAAACTCAAGTGAAATTTGTAATGACTTCAATTTAAT	600
Db	541	ACTAAATTTAATTTTACAAACCAAACTCAAGTGAAATTTGTAATGACTTCAATTTAAT	600
QY	601	GGCCCAATCTTAAGATTAAACCCAGATTGGATTAACCTTATACCTTTCAAGAAAG	660
Db	601	GGCCCAATCTTAAGATTAAACCCAGATTGGATTAACCTTATACCTTTCAAGAAAG	660
QY	661	ATCTTACCAATGACGTCACACTGCAAGTTGTCCTTGGCCAGTAGTACGATTAGTGA	720
Db	661	ATCTTACCAATGACGTCACACTGCAAGTTGTCCTTGGCCAGTAGTACGATTAGTGA	720
QY	721	ACAAATGGTGATGATGGATGTTGATTTGGGGAATGGTCAAAATTACTAATACAGATCT	780
Db	721	ACAAATGGTGATGATGGATGTTGATTTGGGGAATGGTCAAAATTACTAATACAGATCT	780
QY	781	ATTGCTCAAACTAAAAACCCACTGATTAATCAAAATCTTTCAACTTTTAATTCAGAGAGA	840
Db	781	ATTGCTCAAACTAAAAACCCACTGATTAATCAAAATCTTTCAACTTTTAATTCAGAGAGA	840
QY	841	ATGCCGGGGCAAAACAAATAGATAGATCTTCAATGTGAATGTCAAGCATAGAAATTTAAAGA	900
Db	841	ATGCCGGGGCAAAACAAATAGATAGATCTTCAATGTGAATGTCAAGCATAGAAATTTAAAGA	900
QY	901	TCCTTCCAAATTAGATGAATAAATTTGTTTATCCAGAAATGAGCTGGTCTGGAAGGAATPAA	960
Db	901	TCCTTCCAAATTAGATGAATAAATTTGTTTATCCAGAAATGAGCTGGTCTGGAAGGAATPAA	960
QY	961	AATATTAACAAGATTTAGCTACTGGAAGTTTGGCAAGCAACGAAGATTTGATTTCTTGAC	1020
Db	961	AATATTAACAAGATTTAGCTACTGGAAGTTTGGCAAGCAACGAAGATTTGATTTCTTGAC	1020
QY	1021	ATACCCGGGACATCCACAGATTAACCTTAAAGAAAGATTCAGTTAAGCTATTTTCAAGCTA	1080
Db	1021	ATACCCGGGACATCCACAGATTAACCTTAAAGAAAGATTCAGTTAAGCTATTTTCAAGCTA	1080
QY	1081	TACTTAAACCTCAGATTAAATCTTTATCAATTCATTTGGTGAATTAATTTTGGTAC	1140
Db	1081	TACTTAAACCTCAGATTAAATCTTTATCAATTCATTTGGTGAATTAATTTTGGTAC	1140
QY	1141	TCGGAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1200
Db	1141	TCGGAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1200
QY	1201	TTGGAATCAAGTTAAAAACATGATTAATGGAAGCTTCAAGCATGATTAACGATCAACAA	1260
Db	1201	TTGGAATCAAGTTAAAAACATGATTAATGGAAGCTTCAAGCATGATTAACGATCAACAA	1260
QY	1261	AAGGGAACAAGCAACAGAGTGAATAGTAGGCTCAACAGGAGCTGGAAACAGGAAT	1320
Db	1261	AAGGGAACAAGCAACAGAGTGAATAGTAGGCTCAACAGGAGCTGGAAACAGGAAT	1320

QY	1321	ACACTAGAACCTTCTCAAAACAGTTTCTAATCCCTACTTAAATCTCTATGCTAGTTTGG	1380
Db	1321	ACTACTAACACTTCTCAAAACAGTTTCTAATCCCTACTTAAATCTCTATGCTAGTTTGG	1380
QY	1381	ATTGATGATGAACCAACTCTCTCAACCAAAATAGATGAACATAATTGGGAGATCTCTAC	1440
Db	1381	ATTGATGATGAACCAACTCTCTCAACCAAAATAGATGAACATAATTGGGAGATCTCTAC	1440
QY	1441	GTTATTGGAACGAAATATATGCTGATATACAGATTAGTATTCAAAATGAAATTTCCAAAT	1500
Db	1441	GTTATTGGAACGAAATATATGCTGATATACAGATTAGTATTCAAAATGAAATTTCCAAAT	1500
QY	1501	ACTAATGCAGAGAAACCTTTATGCGAAACAAATGGTGTGTTGGTTTACTTCAACAGGT	1560
Db	1501	ACTAATGCAGAGAAACCTTTATGCGAAACAAATGGTGTGTTGGTTTACTTCAACAGGT	1560
QY	1561	TCAAGAGTAGTATTAAAGCTTCTTTAAAGCGTGATCAAGCTCCACACTGGAAACTTCCAA	1620
Db	1561	TCAAGAGTAGTATTAAAGCGTCTTTTAAAGCGTGATCAAGCGTCCAACTGGAAACTTCCAA	1620
QY	1621	CCTTTCTTATACGATTGGTTATTATAGATACCAACAACTAGAACAGAACCTTTCTGG	1680
Db	1621	CCTTTCTTATACGATTGGTTATTATAGATACCAACAACTAGAACAGAACCTTTCTGG	1680
QY	1681	TACGGAACATATTAAGCTTTTAAACAACAGCCCTTACAGCGTATTAGATTCTCCAAAGATA	1740
Db	1681	TACGGAACATATTAAGCTTTTAAACAACAGCCCTTACAGCGTATTAGATTCTCCAAAGATA	1740
QY	1741	GGTACTGAACCAATCAATTATTAAGAAACCTCAATTACCTGGTTATGGGTGAGAT	1800
Db	1741	GGTACTGAACCAATCAATTATTAAGAAACCTCAATTACCTGGTTATGGGTGAGAT	1800
QY	1801	CTAATCTGAAGAAGGCTGACGAAGTTTCTTATATCTCATATATTAAGACACACAGGTGAC	1860
Db	1801	CTAATCTGAAGAAGGCTGACGAAGTTTCTTATATCTCATATATTAAGACACACAGGTGAC	1860
QY	1861	ACACCAGAAAGCCGGAAGCATCTTCCAACTGCGTATTTCTGATATATCTTATAGATACAT	1920
Db	1861	ACACCAGAAAGCCGGAAGCATCTTCCAACTGCGTATTTCTGATATATCTTATAGATACAT	1920
QY	1921	CATACAGTTTATAGATTTGATGGAATTAGAAATTAACCTTAATGTTGGGTTAAACATCA	1980
Db	1921	CATACAGTTTATAGATTTGATGGAATTAGAAATTAACCTTAATGTTGGGTTAAACATCA	1980
QY	1981	AGCTTCTTAACTCAATATGACCAATATCCAAACGCTGTAGAAATGATTTCTCTGCACAAACA	2040
Db	1981	AGCTTCTTAACTCAATATGACCAATATCCAAACGCTGTAGAAATGATTTCTCTGCACAAACA	2040
QY	2041	TACTTAAATTCACAATTTGATTTGATGCTAGAACATCTGATTTACCAACCAACACCATTC	2100
Db	2041	TACTTAAATTCACAATTTGATTTGATGCTAGAACATCTGATTTACCAACCAACACCATTC	2100
QY	2101	GGAAACAACACCAAGTTATTTTACGATACACTGGGTATACGTTTCTCATCAATTAAGAT	2160
Db	2101	GGAAACAACACCAAGTTATTTTACGATACACTGGGTATACGTTTCTCATCAATTAAGAT	2160
QY	2161	ATTAGAACAACTCTCCCTGGTATCCAGTTATGTAATCTTCTTATCCAAATATGAATTAAT	2220
Db	2161	ATTAGAACAACTCTCCCTGGTATCCAGTTATGTAATCTTCTTATCCAAATATGAATTAAT	2220
QY	2221	AAATCTAGTATTATACATTAAGATTAGCTGATCACTCAAGTAACCTGATGGCTCAAGCTCA	2280
Db	2221	AAATCTAGTATTATACATTAAGATTAGCTGATCACTCAAGTAACCTGATGGCTCAAGCTCA	2280
QY	2281	TTCACTGTCACAAGTTTATATGACGTTAATGAAATTTGGTAAATCTTACCTTTATTAAGAC	2340
Db	2281	TTCACTGTCACAAGTTTATATGACGTTAATGAAATTTGGTAAATCTTACCTTTATTAAGAC	2340
QY	2341	AATTCATCTTATACAGTAATGCTGTGTATGTTGCAATTTGTTCTCATCAAAACCTGCT	2400
Db	2341	AATTCATCTTATACAGTAATGCTGTGTATGTTGCAATTTGTTCTCATCAAAACCTGCT	2400
QY	2401	TCTCTCTGATCATATACCTGCTGTAATATACATTTAATCAAGAACTTATCTGATTTGCTTT	2460

```

|||||
Db 2401 TCTCTGATCATATCTGCTGTAATATACATTTAATCAGAACTTATCTGATTTGCTTT 2460
QY 2461 GAAGGTTCTGGTCTAAGTATATCATCTGATTTCTGGGGAACAATCCAAATCCAAACCGAT 2520
Db 2461 GAAAGTTCTGGTCTAAGTATATCATCTGATTTCTGGGGAACAATCCAAATCCAAACCGAT 2520
QY 2521 GAGTACTTAATTCAAAATGGGTTCACTAGTCAAGTGGCTGGAAGTCTGTTTACAAACCA 2580
Db 2521 GAGTACTTAATTCAAAATGGGTTCACTAGTCAAGTGGCTGGAAGTCTGTTTACAAACCA 2580
QY 2581 AGCTTCTTAAACAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 2640
Db 2581 AGCTTCTTAAACAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 2640
QY 2641 GTGGTGTGATCCCTGATATGATTTTAAACAAACCAACCTCTTAAAGTTCAGATCCAA 2700
Db 2641 GTGGTGTGATCCCTGATATGATTTTAAACAAACCAACCTCTTAAAGTTCAGATCCAA 2700
QY 2701 TACTTATGATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
Db 2701 TACTTATGATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
QY 2761 TATTAACAACCTTGGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 2820
Db 2761 TATTAACAACCTTGGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 2820
QY 2821 GGTATCTCTGCAATTTATGATCATCTGATGATGATGATGATGATGATGATGATGATGAT 2880
Db 2821 GGTATCTCTGCAATTTATGATCATCTGATGATGATGATGATGATGATGATGATGATGAT 2880
QY 2881 AGAAAAATTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGATGATGAT 2940
Db 2881 AGAAAAATTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGATGATGAT 2940
QY 2941 GCGTGTGTTCAATTTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGAT 3000
Db 2941 GCGTGTGTTCAATTTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGAT 3000
QY 3001 GCGTGTGTTCAATTTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGAT 3060
Db 3001 GCGTGTGTTCAATTTACAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGATGATGAT 3060
QY 3061 GCTCAGCTTAAACCATCTGACCAAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGAT 3120
Db 3061 GCTCAGCTTAAACCATCTGACCAAAAGAGGTTCAAAACCAACATTCAAAAAGTTGATGAT 3120
QY 3121 TCTGTGTGCTTACAAACCAACATCTGCTTAAACCAACATTCAAAAAGTTGATGATGAT 3180
Db 3121 TCTGTGTGCTTACAAACCAACATCTGCTTAAACCAACATTCAAAAAGTTGATGATGAT 3180
QY 3181 AAGAAATTA 3189
Db 3181 AAGAAATTA 3189

```

RESULT 2
US-10-131-591A-2

```

; Sequence 2, Application US/10131591A
; Publication No. US20030059799A1
; GENERAL INFORMATION:
; APPLICANT: Nippon Zeon Co., Ltd.
; TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
; FILE REFERENCE: J209
; CURRENT APPLICATION NUMBER: US/10/131,591A
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 3189
; TYPE: DNA
; ORGANISM: Mycoplasma gallisepticum
; FEATURE:

```

OTHER INFORMATION: mgc3 gene
US-10-131-591A-2

```

Query Match 100.0%; Score 3189; DB 14; Length 3189;
Best Local Similarly 100.0%; Pred. No. 0;
Matches 3189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATATTTCTTAAAAAAGTTAAAGTATATCATTTAGTGGTATGATGATGATGATGATGAT 60
Db 1 ATGAATATTTCTTAAAAAAGTTAAAGTATATCATTTAGTGGTATGATGATGATGATGATGAT 60
QY 61 GCTCTGTGTTCTGCAACCTTGGCTTAAACCAATCAGATTAAGATTAACGATTAACACGCA 120
Db 61 GCTCTGTGTTCTGCAACCTTGGCTTAAACCAATCAGATTAAGATTAACGATTAACACGCA 120
QY 121 TTAGTTATATCAGACAGACCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 121 TTAGTTATATCAGACAGACCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 181 AATGTTGCTTGTTCATACAGTTCTTAAAGATGATGATGATGATGATGATGATGATGATGAT 240
Db 181 AATGTTGCTTGTTCATACAGTTCTTAAAGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 AATGAACAATTTATCAATTTAGATAGTTTACTTAAACCAATTAATGTTTATGATGATGAT 300
Db 241 AATGAACAATTTATCAATTTAGATAGTTTACTTAAACCAATTAATGTTTATGATGATGAT 300
QY 301 GATGATGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 301 GATGATGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AGATTGTGATCAAGACAAAGACAGATATATGATGATGATGATGATGATGATGATGATGAT 420
Db 361 AGATTGTGATCAAGACAAAGACAGATATATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GTTCATTTTAAAAAAGATTTATCTACTCAATTAATGATGATGATGATGATGATGATGATGAT 480
Db 421 GTTCATTTTAAAAAAGATTTATCTACTCAATTAATGATGATGATGATGATGATGATGATGAT 480
QY 481 AAGTTGTATTTGGTGTGTTGATTAATCCAGCTCAGCTAATTTAGATTTACTGATGATGAT 540
Db 481 AAGTTGTATTTGGTGTGTTGATTAATCCAGCTCAGCTAATTTAGATTTACTGATGATGAT 540
QY 541 ACTAAATTTATTTTAAACCAACCAACTCAAGGTGAAATGTTATGATGATGATGATGATGAT 600
Db 541 ACTAAATTTATTTTAAACCAACCAACTCAAGGTGAAATGTTATGATGATGATGATGATGAT 600
QY 601 GCGCAATCTTACCTAAAGATTTACACCAAGATTTGATTAATCAATTCAGAAAG 660
Db 601 GCGCAATCTTACCTAAAGATTTACACCAAGATTTGATTAATCAATTCAGAAAG 660
QY 661 ATCTTACCAATGACGTCACACGCTGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 661 ATCTTACCAATGACGTCACACGCTGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 ACAAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 721 ACAAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 ATGCTCAAACTAAACCACTACTGATTAATCAAAATCTTCACTTTTAAATTCAGAGCA 840
Db 781 ATGCTCAAACTAAACCACTACTGATTAATCAAAATCTTCACTTTTAAATTCAGAGCA 840
QY 841 ATGCTGTGTCGAAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 ATGCTGTGTCGAAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 TCTTTCGAATTTAGTGAATAATTTGTTATTCAGAAATGACTGTTCTGAAGAAATTA 960
Db 901 TCTTTCGAATTTAGTGAATAATTTGTTATTCAGAAATGACTGTTCTGAAGAAATTA 960
QY 961 AATATTCAAGATTTAGTGAATAATTTGTTATTCAGAAATGACTGTTCTGAAGAAATTA 1020
Db 961 AATATTCAAGATTTAGTGAATAATTTGTTATTCAGAAATGACTGTTCTGAAGAAATTA 1020

```


1681 TACGAGACATTAACCTTTTAAACACAGCCCTTACAGACGTATTAGATTCGCAAGGTA 1740
|||||
5458 TACGAGACATTAACCTTTTAAACACAGCCCTTACAGACGTATTAGATTCGCAAGGTA 5517
|||||
1741 GGTACTGAAGCAATTAATTAAGAGCACTTCAATACATACCTGTTATGGGTGATAT 1800
|||||
5518 GGTACTGAAGCAATTAATTAAGAGCACTTCAATACATACCTGTTATGGGTGATAT 5577
|||||
1801 CTAACTGAAGAGGTGCTAGAAAGTTCTTAATCTCATATATTAAGGCAAGGTGAC 1860
|||||
5578 CTAACTGAAGAGGTGCTAGAAAGTTCTTAATCTCATATATTAAGGCAAGGTGAC 5637
|||||
1861 ACACGGAAGGCGAAGCATCTTCAATCTGCTATTCGATTAATCTTATGAGTACAT 1920
|||||
5638 ACACGGAAGGCGAAGCATCTTCAATCTGCTATTCGATTAATCTTATGAGTACAT 5697
|||||
1921 CAATCAAGTTTGAAGTTTGAAGATTAAGAAATTAATCTTAAATCTGTTGAGGTTAAAGCATCA 1980
|||||
5698 CAATCAAGTTTGAAGTTTGAAGATTAAGAAATTAATCTTAAATCTGTTGAGGTTAAAGCATCA 5757
|||||
1981 AGCTCTTAACCTCAATTAAGACCAATCCAAACGGCTGAGAAATGATTCGCAACANCA 2040
|||||
5758 AGCTCTTAACCTCAATTAAGACCAATCCAAACGGCTGAGAAATGATTCGCAACANCA 5817
|||||
2041 TACTTAAGATCACAAATGGATTAGCTAGACATCTGATTTACCAACCAACAGACATTC 2100
|||||
5818 TACTTAAGATCACAAATGGATTAGCTAGACATCTGATTTACCAACCAACAGACATTC 5877
|||||
2101 GGAACACACACCAAGTTATTCAGTATCAGTCTGATTCATCAATTAAGAT 2160
|||||
5878 GGAACACACACCAAGTTATTCAGTATCAGTCTGATTCATCAATTAAGAT 5937
|||||
2161 ATTAGAACCAATCTTCCCTGGTACCAAGTATGCTACTTCTTATTCACAAATGGAATAT 2220
|||||
5938 ATTAGAACCAATCTTCCCTGGTACCAAGTATGCTACTTCTTATTCACAAATGGAATAT 5997
|||||
2221 AAATCTAGCTTTTATTAATTAAGATTAAGTCTGACCTCAAGTAAACCTGATGCTCAAGTCA 2280
|||||
5998 AAATCTAGCTTTTATTAATTAAGATTAAGTCTGACCTCAAGTAAACCTGATGCTCAAGTCA 6057
|||||
2281 TTACAGTCCACAAAGTTTATTAAGATTAAGTCTGACCTCAAGTAAACCTGATGCTCAAGTCA 2340
|||||
6058 TTACAGTCCACAAAGTTTATTAAGATTAAGTCTGACCTCAAGTAAACCTGATGCTCAAGTCA 6117
|||||
2341 AATTCATCTTACAGTAAATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2400
|||||
6118 AATTCATCTTACAGTAAATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6177
|||||
2401 TCTCTGATCATATATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2460
|||||
6178 TCTCTGATCATATATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6237
|||||
2461 GAAGGTTGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2520
|||||
6238 GAAGGTTGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6297
|||||
2521 GAGTACTTAATCAAAATGGGTTCACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2580
|||||
6298 GAGTACTTAATCAAAATGGGTTCACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6357
|||||
2581 AGCTCTTAATCAAAAGTTTGAAGTCACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2640
|||||
6358 AGCTCTTAATCAAAAGTTTGAAGTCACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6417
|||||
2641 GTGGTGTGATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2700
|||||
6418 GTGGTGTGATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6477
|||||
2701 TACTTAAGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2760
|||||
6478 TACTTAAGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6537
|||||
2761 TATTAACAATTTGGCGCTTACCTCTACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2820
|||||

|||||
6538 TATTAACAATTTGGCGCTTACCTCTACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6597
|||||
2821 GGTATTTCTGCAATTAATGATCATCTTACGATTAAGTATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2880
|||||
6598 GGTATTTCTGCAATTAATGATCATCTTACGATTAAGTATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6657
|||||
2881 AGAAATTTCAAGCAAGGTTCAAAACCAATCAAAAAAGTTGATACCTGATGCT 2940
|||||
6658 AGAAATTTCAAGCAAGGTTCAAAACCAATCAAAAAAGTTGATACCTGATGCT 6717
|||||
2941 GCTGTTGCTTCAAGTATTAAGAAATTTATACCAAACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 3000
|||||
6718 GCTGTTGCTTCAAGTATTAAGAAATTTATACCAAACTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6777
|||||
3001 GCTGCTTTAGGCTGCTGATTAATCTGCTGATTAAGAAACCTGCTGCTGCTGATTAACCTGCT 3060
|||||
6778 GCTGCTTTAGGCTGCTGATTAATCTGCTGATTAAGAAACCTGCTGCTGCTGATTAACCTGCT 6837
|||||
3061 GCTCAGCTTAACCAATCTGCTGATTAAGAAACCTGCTGCTGCTGATTAACCAACCTGCTGCTGCTGATTAAC 3120
|||||
6838 GCTCAGCTTAACCAATCTGCTGATTAAGAAACCTGCTGCTGCTGATTAACCAACCTGCTGCTGCTGATTAAC 6897
|||||
3121 TCTGCTGCTGCTGCTGATTAAGAAACCTGCTGCTGCTGATTAAGAAACCTGCTGCTGCTGATTAAGAAACCT 3180
|||||
6898 TCTGCTGCTGCTGCTGATTAAGAAACCTGCTGCTGCTGATTAAGAAACCTGCTGCTGCTGATTAAGAAACCT 6957
|||||
3181 AAAGATTA 3189
|||||
6958 AAAGATTA 6966
|||||

RESULT 4
US-10-131-591A-79
; Sequence 79, Application US/10131591A
; Publication No. US20030059799A1
; GENERAL INFORMATION:
; APPLICANT: Nippon Zeon Co., Ltd.,
; TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
; FILE REFERENCE: J209
; CURRENT APPLICATION NUMBER: US/10/31.591A
; FILING DATE: 2002-08-15
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 3129
; TYPE: DNA
; ORGANISM: Mycoplasma gallisepticum
; FEATURE:
; OTHER INFORMATION: Modified mgc3 gene
US-10-131-591A-79

Query Match 95.9%; Score 3057.8; DB 14; Length 3129;
Best local similarity 98.7%; Pred. No. 0; Indels 0; Gaps 0;
Matches 3083; Conservative

65 TTGGTCTGCAAGCTTTGGCTTAAAGCAATGATTAAGAGTAAACGATTAACGCAATTTAG 124
|||||
5 TCGGTTCTGCAAGCTTTGGCTTAAAGCAATGATTAAGAGTAAACGATTAACGCAATTTAG 64
|||||
125 TTAATCAAGCAAGCAAGCTGATGCTAATTTGTTAGACTTCCAGCTCTTGGACAAAATG 184
|||||
65 TTAATCAAGCAAGCAAGCTGATGCTAATTTGTTAGACTTCCAGCTCTTGGACAAAATG 124
|||||
185 GTTGGTCTGCAAGCTTTGGCTTAAAGCAATGATTAAGAGTAAACGATTAACGCAATTTAG 244
|||||
125 GTTGGTCTGCAAGCTTTGGCTTAAAGCAATGATTAAGAGTAAACGATTAACGCAATTTAG 184
|||||
245 GAACAAATTTCAAAATTTGATGCTTAAACCAATTAATGCTTAAATTTAGTCAATGATG 304
|||||
185 GAACAAATTTCAAAATTTGATGCTTAAACCAATTAATGCTTAAATTTAGTCAATGATG 244
|||||
305 ATTGGTGTGATCAAGCAAGTAAACCAATTAATGCTTAAATTTAGTCAATGATGAT 364
|||||


```

RESULT 5
US-10-205-220-1
: Sequence 1, Application US/10205220
: Publication No. US20030170663A1
: GENERAL INFORMATION :
: APPLICANT: Fraser et al.
: TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
: TITLE OF INVENTION: Thereof, and Uses Thereof
: FILE REFERENCE: PB193P1D1
: CURRENT APPLICATION NUMBER: US/10/205,220
: CURRENT FILING DATE: 2002-07-26
: PRIOR APPLICATION NUMBER: US 08/545,528
: PRIOR FILING DATE: 1995-10-19
: PRIOR APPLICATION NUMBER: US 08/488,018
: PRIOR FILING DATE: 1995-06-07
: PRIOR APPLICATION NUMBER: US 08/473,545
: PRIOR FILING DATE: 1995-06-07
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 580073
: TYPE: DNA
: ORGANISM: Mycoplasma genitalium

```

QY	2770	TTTGGGCGCTTACCTTCATCGGTAGTGGCTACAGCAAAATGGGTACTAATTAAGTAATTCCTT	2829
Db	228703	TTTGCAGCACATCCCTCCATGGGATGCCCTGTATCCATAGAGTTCCTAGTGGGATCTTG	228762
QY	2830	GCAATTTATGATCATCTTTAGGATTTAGCTATCGTATTCCTTTAAGAGCTCAAAAGAAATTA	2889
Db	228763	TTTATCTGTGTACTGTATGAGACTGTGGATTTGGATCCCAATGTACAGGGATACAAAATC	228822
QY	2890	CAACACAAAGAGTTCAAAACACATTCACAAAAAAGTTGATACCTTGACTGCTGCTGTGCT	2949
Db	228823	CAAAATGCATTCGTTTGTATATGCTTTTAAAAAGTTGATATACACTCACAACGTGCTGTGGT	228882
QY	2950	TCAGTTTACAAGAAGATTATTACCCAAACCTGCGTAACGTAACAAAAAAACCTGCTGCTTTA	3009
Db	228883	AGTGTGTAACAAAAAGATTATTATACCCAATCGGTGTGTGTAAAAAAGCACACTAGTGCATTTG	228942
QY	3010	GGTGCTGTAAATCTGGTGTATAGAAACCTGCTGCTGCTGCATTAACCTGCGTCGTCACAC	3068
Db	228943	AAAGTGTGTAATCTCACTAGTGTAAAAAAACCTGCTGCTTTTATTAACCACTGTTTCAACC	229001

APPLICANT: ZWIBBEL, LAURENCE J.
 TITLE OF INVENTION: MOSQUITO OLFATORY GENES, POLYPEPTIDES, AND METHODS OF
 TITLE OF INVENTION: USE THEREOF
 FILE REFERENCE: N7841
 CURRENT APPLICATION NUMBER: US/10/056,405
 CURRENT FILING DATE: 2002-01-24
 PRIOR APPLICATION NUMBER: 60/264,649
 PRIOR FILING DATE: 2001-01-26
 NUMBER OF SEQ ID NOS: 23
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 10
 LENGTH: 4985
 TYPE: DNA
 ORGANISM: Anopheles gambiae
 US-10-056-405-10

[illegible]

Db	2770	ATATATATTAATAAATCATAATATAAATTTATTATTATGATATAATATGATATATATGAT	2711
Qy	611	TAC-CTAAGATTTTACACCCAGATTGGTATPACATTAPACATTCGAAGAAGATCTTACCA	669
Db	2710	AACAT	2651
Qy	670	AATPACGTCACACACTGCAGTTGCTTCCTGGCCACTAGGTAGAGTTAGTGGACAATGCT	729
Db	2650	TATATTGTTATATAATTAATAAATAAATAATTAATAATATATATATATATATATATATAT	2591
Qy	730	GATGATGGAGCTTTGATTTGCGGAATGTCOAATACTATPACAGATCCATATGCTCA	789
Db	2590	AATATATATATAATTAATAATTAATAATPATAATATATATATATATATATATATATATATAA	2531
Qy	790	ACTAAACCACTACTGATTAATCAAAATCCCTCAACTTTTATTCAGAGCAATGCGCTGT	849
Db	2530	AGTATATATTTTATTAATAATATAGTAATATATATATATATATATATATATATATATATAT	2471
Qy	850	GCAACATATAGATGATTCCTCAATTGATGTCACGATAGATAATTAACAACCTCTTCCAA	909
Db	2470	AAT	2411
Qy	910	TTAGATGAAAAATTTGTTTATCCAGAAATGCACTGGTCTGGAAGAATAAATAATTTACA	969
Db	2410	AAT	2351
Qy	970	AGATTAGTACT 981	
Db	2350	AAATTCGTATTT 2339	

```

RESULT 7
US-10-094-240-10/C
: Sequence 10, Application US/10094240
: Publication No. US20030082637A1
: GENERAL INFORMATION:
: APPLICANT: ZWIEBEL, LAURENCE J.
: TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
: FILE REFERENCE: M8289
: CURRENT APPLICATION NUMBER: US/10/094,240
: PRIOR FILING DATE: 2001-03-08
: PRIOR APPLICATION NUMBER: 10/056,405
: PRIOR FILING DATE: 2002-01-24
: PRIOR APPLICATION NUMBER: 60/264,649
: PRIOR FILING DATE: 2001-01-26
: NUMBER OF SEQ ID NOS: 27
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10
: LENGTH: 4985
: TYPE: DNA
: ORGANISM: Anopheles gambiae
US-10-094-240-10

```

Query Match	2.0%;	Score 62.6;	DB 14;	Length 4985;
Best Local Similarity	45.5%;	Pred. NO. 0.015;		
Matches 333; Conservative	0;	Mismatches 396;	Indels 3;	Gaps 3;

Oy	253	TTCAAATTAGTACTTTTTAGCAAAACCTTAATATGGTTTAGATCTAAGGATGATGTGGCT	312
Db	3070	ATTAAAAAAAACAATAATTAATAAACATTAATAAGAATTAACAACAACAACAATAATTAAGAA	3011
Oy	313	GGATACAAAGTAAGAACAAATAGTTTCA-GATTACACAACATGCAGAAATAGATTGGATCA	371
Db	3010	CACACACACACACATATATAGATTAATTAATTAACATATGCATATTAAGATTAATTA	2951
Oy	372	AAGACAACAAGACCATTTATVTCCTCGTTGGTTAATGATGAAGCTAAGCTTCAAT-TTAA	430
Db	2950	TAACAATTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2891
Oy	431	AAAGAATTAATACATCAACTAATGAAATGGTATTAAGAAACAACAATTCTAAGTTTGAA	490
Db	2890	AAAAAATTAATTAATTAAGATAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2831

Qy	491	TTGGTGGTGGATAAATCCAGCTACGTAATATAGATTACTGATGATGGGACTAAATTTA	550
Db	2830	ATAATTAATTAATAAATAAATAATTAATTAATTAAGATAATAATAAATAAATAATTAATA	2771
Qy	551	ATTATTACAACCAACAGCTCAGGCGAATAATGTGATGACTCATTTTGGATCGCCAAATCT	610
Db	2770	ATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2711
Qy	611	TAC-CTAAGATTTTACACCCAGATTTGGTATTAACCTTAATACATTCAACAGAAAGCTTTACA	669
Db	2710	AACATATAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2651
Qy	670	AATGACGTCAACTGCAGTTGTTCTCTGGCCAGTAGGTAGCTTACTGCAACAATGCT	729
Db	2650	TATTAATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2591
Qy	730	GATCATGGGATGTTTGATTGTGGGAATGGTCAAAATACTAATACAGATCCATTGCTCAA	789
Db	2590	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2531
Qy	790	ACTAAACACGACTGATGAATCAAAATCCCTCAACTTTTAATTCAGAGCAATGCCGTG	849
Db	2530	AGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2471
Qy	850	GCAACAATAGATAGATTCGATTCATTTGAGATGTCAGCAGTAGAATTAACCACTTTCCAA	909
Db	2470	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2411
Qy	910	TTAGATGAAAAATTTGTTTATCCGAATGAGCTGGTCTGAGAGGAATAAAAATATTCA	969
Db	2410	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2351
Qy	970	AGATTAGCTACT 981	
Db	2350	AAATTCGTGTAAT 2339	

```

RESULT 8
US-10-311-455-1798/C
; Sequence 1798, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1798
; LENGTH: 8771
; TYPE: DNA
; ORGANISM: Official Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1798

```

Query Match	1.9%	Score 60.8	DB 12	Length 8771
Best Local Similarity	46.3%	Pred. NO 0.047		
Matches	200	Conservative	0	Mismatches 232
			Indels	0
			Gaps	0

Oy	256	AAATTAGACTTTTACTAAACCATATATATGCTTTAGATCTAAGTGATGATGGCGGA	315
db	5053	ATATCTATATACATCAATAAAAATTTATATCAACGCTTTAAAAAAAACCTTTAACTTA	4994

Query Match	Similarity	1.9%	Score 60.2	DB 12	Length 3489
Best Local	Similarity	61.6%	Pred. No. 0.04		
Matches 114	Conservative	0	Mismatches 68	Indels 3	Gaps 1
Oy	2998	CCGCTGCTTTAAGTGGCTGCTGTAATCTGCTATTAAGAAACCTGCGCGCTGCTAAACCT	3057		
Db	2236	CATCTGCTGCTGCTCTCTCATCTCTGCTGCTGCTCAACCTGCTGCTGCTCATCCT	2177		
Oy	3058	GCTGCGTCAGCTAAACCATCTGCTACACCAAAAGCTAGCTACCAAGCTAAACCACTGGCG--	3115		
Db	2176	GCTGCTGCTGCTCATCTCTGCTGCTGCTATCTCTGCTGCTGCTCACTCCGCTGCTCTCT	2117		
Oy	3116	-CTAAATCTGCTGCGCTTACAAAACCAACTGCTCTTAAGCCGCTGCTGCCAAAACCAAC	3174		
Db	2116	GCTCATCTGCTGCTCTGCTCATCTCTGCTGCTGCTGCTCATCTGCTGCTGCTCATCTCTCT	2057		
Oy	3175	GCTCC 3179			
Db	2056	GCTGC 2052			

QY	241	ATATGACAACTTATATCAAAATAGATAGTTTACTTAACCACTTATATGATGATTCAGATCAGG	525106
Db	525047	ATTTAAAAAATAAACACAGAAATTTATATACATACATCTTCATTAAGTAAATTTTATT	525106
QY	301	GATGATTGTGGTGATACAAAGTAAACCAATAGTTTCAGATTACACAACCTGACGAAT	360
Db	525107	TTAAAAAAGAAGAAAAAATTTATTAACCACTTATCTCAAAAAGAAATTAATAAAAA	525166
QY	361	AGATTGATCAAGACAAAGACAGATTTATGCTGTGGTTAATGATGACCTAAC	420
Db	525167	GCATTTGTTAAATTAATCTCTTCATCTTTAAAAAATAATTAACAATCAAGACACACA	525226
QY	421	GTTTCATTTAAAAAGAAATTAATACTAATCTCAATAGAAATGGTATATGAACCAACATCT	480
Db	525227	ATTTATTTCCAAAAAATCCACAGATATTTTATATTCAGAAAAATTTAAATTAATTT	525286
QY	481	AACTTTGTAATTTGGTGTGTATATCCGCTCAGCTATATGATTTACTGATGG	540
Db	525287	GTTGAATTAATAAACTGATTAATCTTTAAATATCAATGGCAAAATTAAGAGATTTTGATGG	525346
QY	541	ACTAAATTTATTTTACAAACCAACTCAAGGTGAATTTGTTAATGACTTCATTTAGAT	600
Db	525347	TATTTAAGAAATTTTAC - ACAATACCACTAAGAAAAAAGAGATTAATGATTTATTC	525404
QY	601	GGCGCAATCTTACCTAAAGATTTACACCCAGATTTGGTATACCTTATCATTTCAAGAAAG	660
Db	525405	AAGCAAAAATTAACAACAAGCCATATGATATTTATCTACATTTGCATTAATACACCAAG	525464
QY	661	ATCTTACCAATGACGTCAACA	682
Db	525465	ATTTTCAAAATAGCTCAGA	525486

```

RESULT 11
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
;
GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des
;
FILE REFERENCE: E01/1208/MO
;
CURRENT APPLICATION NUMBER: US/10/312,841
;
CURRENT FILING DATE: 2002-12-30
;
NUMBER OF SEQ ID NOS: 2
;
SEQ ID NO 1
;
LENGTH: 3673778
;

```

TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (3294164)
US-10-312-841-1

Query Match 1.7%; Score 53.6; DB 12; Length 3673778;
Best Local Similarity 47.1%; Pred. No. 29;
Matches 196; Conservative 0; Mismatches 219; Indels 1; Gaps 1;

QY 801 TACTGATATCAAAATCCTTCACCTTTAATTCAGAGCAATGCGTGCAGCAATAG 860
DB 98780 TAATATCCAAATATTCACATCACCATTAAATTAACCTCTCTCTCATTAAATTA 98721
QY 861 ATACGATCTCAATGTAATGTCAGCATAGATAATTAACATCTTCCAAATAGATAAA 920
DB 98720 CTTTATACCTTTATATAAATCAATTAACATATTTACTCATTTTAAATTAATTTCTC 98661
QY 921 ATTTGTTATCCGAATGCGCTGTTCTGAAGCAATAAATATTACAAGATTAGCTAC 980
DB 98660 ATCTTCTACATTAATTAATTAACATTTTCAATATTTCTAAATACATATTTCTTCATCACA 98601
QY 981 TGGAACTTTGCCAAGCAAGCAAGATTTGGATTCTTGACATACCGGAGCTCCACAGT 1040
DB 98600 AATATATATTAATTAATTAATTAATTAATTTCTCTCAATCTACGACTTATCTTTCATTT 98541
QY 1041 TACTTTAAAGAAATTCAGTTAAGCTATTTTCAAGACTTAACTTAACTCAAGTTATTC 1100
DB 98540 TTTCTCAATATTAATTTAAATTAATTAATTTAAATTTAACTTAACCAATTTA-TC 98482
QY 1101 TTTATCATTCATGGATGATGATTTATATTTTGGTACTCGAATTAACATCATATG 1160
DB 98481 TCTCAGCTTTTTCATCTTAATTAATCTATTTTAAATTTTAAATTAATTAATTTTACC 98422
QY 1161 GTACTATTCATTCCTCACTAGATTAATCTGATCTAACCGTTTGAATCAAGTTAAA 1216
DB 98421 TAACCTCAATTAATTAATTAATTTTCTCTATATTTCTTTTAAATTTTAA 98366

RESULT 12
US-10-349-680-148
Sequence 148, Application US/10349680
Publication No. US20030176654A1
GENERAL INFORMATION:
APPLICANT: Cassell, Gail
APPLICANT: Chen, Ellison
APPLICANT: Glass, Jennifer
APPLICANT: Glass, John
APPLICANT: Helner, Cheryl
APPLICANT: Lefkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
FILE REFERENCE: UAB-13403/22
CURRENT APPLICATION NUMBER: US/10/349,680
PRIORITY FILING DATE: 2003-01-23
PRIORITY APPLICATION NUMBER: US 09/601,198
PRIORITY FILING DATE: 2000-12-08
PRIORITY APPLICATION NUMBER: PCT/US99/01972
PRIORITY FILING DATE: 1999-01-29
PRIORITY APPLICATION NUMBER: US 60/073,189
PRIORITY FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn version 3.2
SEQ ID NO 148
LENGTH: 3057
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-10-349-680-148

Query Match 1.6%; Score 52.4; DB 12; Length 3057;

Best Local Similarity 45.6%; Pred. No. 1.5;
Matches 220; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

QY 754 AATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 813
DB 1175 AATTCGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1234
QY 814 AATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 873
DB 1235 TATTCGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1294
QY 874 TTGAATGTCAGCATAGATAATTAACATCTTCCAAATAGATAATTAATTAATTAATTAATTA 933
DB 1295 CTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1354
QY 934 GAATGAGCTGTTCTGAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 993
DB 1355 AATTTGAATTTAGCAATTTGATTTTAACTTTGATTTGATTTGATTTGATTTGATTTGATTTG 1414
QY 994 AGCAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1053
DB 1415 GCATACCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1474
QY 1054 GATTCAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1113
DB 1475 ATTCCTCAATTTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
QY 1114 GGTGATAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1172
DB 1535 GGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1594
QY 1173 CCCAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1232
DB 1595 CTATTTCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1654
QY 1233 TT 1234
DB 1655 TT 1656

RESULT 13
US-10-091-007-55
Sequence 55, Application US/10091007
Publication No. US20030170782A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Le Page, Richard W F Wells, Jeremy M
APPLICANT: Hanniffy, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PMC/P21978W0
CURRENT APPLICATION NUMBER: US/10/091,007
PRIORITY FILING DATE: 2002-03-06
PRIORITY APPLICATION NUMBER: GB 9921125.2
PRIORITY FILING DATE: 1999-09-07
NUMBER OF SEQ ID NOS: 276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 55
LENGTH: 2451
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-091-007-55

Query Match 1.6%; Score 51.6; DB 12; Length 2451;
Best Local Similarity 45.2%; Pred. No. 1.9;
Matches 189; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 1184 TATCTGATCTAACCGCTTTGAATCAAGTATATGAAGCTTCAAGCACTG 1243
DB 1016 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1075
QY 1244 ATAAAGCTTACAAACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1303
DB 1076 AAAAGGATTTAGTCCGAAAGATTAATCAAGTTTGAAGAAACATTTAGAAGTTTGTGTA 1135

[illegible]

Query Match	Best Local Similarity	1.6%;	Score 50.8;	DB 14;	Length 5314;	
Matches	386;	Conservative	0;	Mismatches 552;	Indels 6;	Gaps 1;
QY	207	TAGAGATGTTGGATGAACCTTATATACAGCAGCTAATGGAACATTTTCAAATTACATG	266			
DB	1380	TAAAAATGATATGATGAGAAAAAGAAAAATATATATATTAACAGGAAGATATAT	1439			
QY	267	TTTTACTAAACCTTTATGTTTGGATCTAAGTAGATGTTGGTGATACAAAGTAA	326			
DB	1440	AAAAATGACCAATGTCTATATATATATATATTAATTCAAATTCATATGATATGACAAAA	1499			
QY	327	ACAAATGTTTCGATTACACAACTACGGAATATGGATTCGATCAAGACCAACAGAGC	386			
DB	1500	TGCATCTACATTTAGATGAACCAATATATAGTAAACCTTTGAAGTTATCTTTATATGT	1559			
QY	387	ATATTATGCTGCTGGTTATGATGATCAAGCTTCACTTTAAAAAGATTAATACTAA	446			

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 06:22:40 : Search time 11360 Seconds

(without alignments)
11484.225 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189

Sequence: 1 atgaatattctcaaaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_pro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3189	100.0	3189	6	AX665165	AX665165 Sequence
2	3145.8	98.6	3189	1	AB023292	AB023292 Mycoplasma
3	3144.2	98.6	301042	1	AE016967	AE016967 Mycoplasma
4	3143.4	98.6	8354	1	AF214004	AF214004 Mycoplasma
5	3143.4	98.6	10651	6	AX113685	AX113685 Sequence
6	3057.8	95.9	3129	6	AX665242	AX665242 Sequence
7	1052.4	33.0	7141	1	AF083976	AF083976 Mycoplasma
8	993.2	31.1	1131	1	AB033210	AB033210 Mycoplasma
9	924.6	29.0	1128	1	AB033211	AB033211 Mycoplasma
10	124.6	3.9	8760	1	MYCWP	M31431 M.genitali
11	124.6	3.9	15787	1	U39698	U39698 Mycoplasma
12	124.6	3.9	110000	6	AR30198_2	Continuation (3 of
13	102.2	3.2	9691	1	MYCATP	M21519 M.pneumonia
14	102.2	3.2	16876	1	AE000002	AE000002 Mycoplasma
15	93	2.9	1618	1	MG034970	U34970 Mycoplasma
16	80.8	2.5	105682	3	AC116957_3	Continuation (4 of
17	78.4	2.5	4695	1	MG034842	U34842 Mycoplasma
18	76.2	2.4	145265	2	AC007926	AC007926 Trypanoso
19	75.2	2.4	155204	2	AC007926	AC007926 Trypanoso
20	75.2	2.4	213544	2	BX510939	BX510939 Danio rer
21	73.6	2.3	151343	5	AL929250	AL929250 Zebrafish
22	72.4	2.3	94434	5	AL929250	AL929250 Zebrafish
23	71.8	2.3	64707	3	AC115607	AC115607 Dictyoste
24	71.8	2.3	225581	2	BX537105	BX537105 Danio rer
25	71.4	2.2	203138	2	BX530070	BX530070 Danio rer
26	70.8	2.2	3502	1	AY212515	AY212515 Mycoplasma
27	69.2	2.2	148871	2	AL929469	AL929469 Danio rer
28	68.8	2.2	266544	3	AC116956	AC116956 Dictyoste
29	68.6	2.2	151802	3	AC114263	AC114263 Dictyoste
30	67.6	2.1	82139	3	AC115684	AC115684 Dictyoste
31	67.4	2.1	12463	3	AC115578	AC115578 Dictyoste
32	67.4	2.1	168244	2	BX537273	BX537273 Danio rer
33	66.6	2.1	111882	3	AC115612	AC115612 Dictyoste
34	66.4	2.1	110000	3	AC116984_3	Continuation (4 of
35	66.4	2.1	180108	2	BX511115	BX511115 Danio rer
36	66	2.1	176267	2	AL929105	AL929105 Danio rer
37	65.8	2.1	250029	3	AE014820	AE014820 Plasmodiu
38	65.4	2.1	169478	2	AC139280	AC139280 Homo sapi
39	65.4	2.1	195532	2	BX004760	BX004760 Danio rer
40	65.4	2.1	250195	3	AE014831	AE014831 Plasmodiu
41	65.2	2.0	158266	2	BX323881	BX323881 Danio rer
42	65.2	2.0	178702	2	BX321884	BX321884 Danio rer
43	65.2	2.0	185107	2	AC118513	AC118513 Rattus no
44	65.2	2.0	243570	2	AC125757	AC125757 Rattus no
45	65	2.0	166583	2	AC137124	AC137124 Mus muscu

ALIGNMENTS

RESULT 1

AX665165

LOCUS AX665165 3189 bp DNA linear PAT 26-MAR-2003

DEFINITION Sequence 2 from Patent EP1275716.

ACCESSION AX665165

VERSION AX665165.1 GI:29290295

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 Okuda,T., Saito,S., Dorsey,K.M. and Tsuzaki,Y.
Modified dna molecule, recombinant containing the same thing, and
uses thereof
Patent: EP 1275716-A 2 15-JAN-2003;

JOURNAL

Zeon Corporation (JP)
Location/Qualifiers
1..3189
/organism="Mycoplasma gallisepticum"
/mol_type="genomic DNA"
/db_xref="taxon:2096"
/note="mgc3 gene"

BASE COUNT 1114 a 576 c 539 g 960 t

Query Match 100.0%; Score 3189; DB 6; Length 3189;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAAATATTTCTAAAAAAGTATTAATGATAGAGGGATGATGATTTGA 60
1 ATGAAATATTTCTAAAAAAGTATTAATGATAGAGGGATGATGATTTGA 60
DB 61 GCTCTGGTTCTGCAAGCTTTGGCTTTAAGCAATGATAGAGATTAACGCA 120
OY 121 TTGATTAATCAAGCAAGAGCGTAGATGCTAATTCGTTAGACTGCAAGTTCGACAA 180
DB 121 TTGATTAATCAAGCAAGAGCGTAGATGCTAATTCGTTAGACTGCAAGTTCGACAA 180
OY 181 AATGGCTGTTGTTCAATACAGTTCTTAGAGATGTTGATTAATTAACAGCAGCT 240
DB 181 AATGGCTGTTGTTCAATACAGTTCTTAGAGATGTTGATTAATTAACAGCAGCT 240
OY 241 AATGGAACAATTAATCAAAATAGATGTTTACTTAACCAATTAATGCTTAATCTAAGT 300
DB 241 AATGGAACAATTAATCAAAATAGATGTTTACTTAACCAATTAATGCTTAATCTAAGT 300
OY 301 GATGATGTTGGTGGATCAAAAGTAAACAATAGTTTCAGATTACCAACTAGCAGAAAT 360
DB 301 GATGATGTTGGTGGATCAAAAGTAAACAATAGTTTCAGATTACCAACTAGCAGAAAT 360
OY 361 AGATTTGATCAAAAGCAAAAGCAATATTATGCTCTGTGGTTAATGATGAAGCTAAC 420
DB 361 AGATTTGATCAAAAGCAAAAGCAATATTATGCTCTGTGGTTAATGATGAAGCTAAC 420
OY 421 GTTCAATTAATAAGAAATTAATCAATCAAAATAGATTGTAATAGAAAACAATTCCT 480
DB 421 GTTCAATTAATAAGAAATTAATCAATCAAAATAGATTGTAATAGAAAACAATTCCT 480
OY 481 AAGTTGTAATGTTGGTGGTGAATATTCACAGCTCAGCTAATTAATTAATGATGATGGG 540
DB 481 AAGTTGTAATGTTGGTGGTGAATATTCACAGCTCAGCTAATTAATTAATGATGATGGG 540
OY 541 ACTAAATTAATTAATTAACAACCAACTCAAGTGAATGTTAATGATTCATTTAGAT 600
DB 541 ACTAAATTAATTAATTAACAACCAACTCAAGTGAATGTTAATGATTCATTTAGAT 600
OY 601 GCGCAATCTTAACCTAAGATTTACACCCAGATTGCTAATTAATTAATTAACAAGAAAG 660
DB 601 GCGCAATCTTAACCTAAGATTTACACCCAGATTGCTAATTAATTAATTAACAAGAAAG 660
OY 661 ATCTTACCAAAATGACGTCACAGCTGATGTTCCCTGSCAGTAGAGATTAAGTGA 720
DB 661 ATCTTACCAAAATGACGTCACAGCTGATGTTCCCTGSCAGTAGAGATTAAGTGA 720
OY 721 ACAAAATGCTGATGATGATGTTGATGTTGGGAATGTCAAATTAATTAACAATCACT 780
DB 721 ACAAAATGCTGATGATGATGTTGATGTTGGGAATGTCAAATTAATTAACAATCACT 780
OY 781 ATTGCTCAAACTAAACCACTACTGATTAATCAAAATCTTCAACTTTTATTCAGAGACA 840
DB 781 ATTGCTCAAACTAAACCACTACTGATTAATCAAAATCTTCAACTTTTATTCAGAGACA 840
OY 841 ATGCGTGTGCAAAACAATAGATAGATTTCAATTAATTAATTAATTAATTAATTAATTA 900
DB 841 ATGCGTGTGCAAAACAATAGATAGATTTCAATTAATTAATTAATTAATTAATTAATTA 900

OY 901 TCTTTCATTAATGATGAAAAATTTGTTATTCAGAAATGACAGTGTCTGAAGAGATATAA 960
DB 901 TCTTTCATTAATGATGAAAAATTTGTTATTCAGAAATGACAGTGTCTGAAGAGATATAA 960
OY 961 AATATTAACAGATTAGTACTGAGAGAGTTTCCAGCAAGCAAGAAATGATGATTCCTGAC 1020
DB 961 AATATTAACAGATTAGTACTGAGAGAGTTTCCAGCAAGCAAGAAATGATGATTCCTGAC 1020
OY 1021 ATACCCGGAGCTCCAAAGTACTTCTTAAGAAAGATCAAGTAAAGATTTTCCAGACTA 1080
DB 1021 ATACCCGGAGCTCCAAAGTACTTCTTAAGAAAGATCAAGTAAAGATTTTCCAGACTA 1080
OY 1081 TACTTAACCTGATTAATTTCTTATCAATTCATGTTGATGATTTTATTTTGGTACC 1140
DB 1081 TACTTAACCTGATTAATTTCTTATCAATTCATGTTGATGATTTTATTTTGGTACC 1140
OY 1141 TCTGAATTAACATATATGATGATTAATTCATTCATTCACATATATGATTAATTCACCGCT 1200
DB 1141 TCTGAATTAACATATATGATGATTAATTCATTCATTCACATATATGATTAATTCACCGCT 1200
OY 1201 TTGAATCAAGTAAACAGATGATATTAAGCTTCCAGCAAGTAAAGATTAACAGTAAACA 1260
DB 1201 TTGAATCAAGTAAACAGATGATATTAAGCTTCCAGCAAGTAAAGATTAACAGTAAACA 1260
OY 1261 AACGGAACAACGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 AACGGAACAACGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
OY 1321 ACTAATTAACCTCTCAACAGATTTCTAATCTTAAATTAATTAATTAATTAATTAATTA 1380
DB 1321 ACTAATTAACCTCTCAACAGATTTCTAATCTTAAATTAATTAATTAATTAATTAATTA 1380
OY 1381 ATTGATTAACCAACTTCTGCAACCAATTAATTAATTAATTAATTAATTAATTAATTA 1440
DB 1381 ATTGATTAACCAACTTCTGCAACCAATTAATTAATTAATTAATTAATTAATTAATTA 1440
OY 1441 GTTATTAACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
DB 1441 GTTATTAACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
OY 1501 ACTAATGAGGAACCTTTATTCGCAACCAATTAATTAATTAATTAATTAATTAATTAAT 1560
DB 1501 ACTAATGAGGAACCTTTATTCGCAACCAATTAATTAATTAATTAATTAATTAATTAAT 1560
OY 1561 TCAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
DB 1561 TCAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
OY 1621 CCTTCTTAATGATGATTTGTTATTTAGATTAATTAATTAATTAATTAATTAATTAAT 1680
DB 1621 CCTTCTTAATGATGATTTGTTATTTAGATTAATTAATTAATTAATTAATTAATTAAT 1680
OY 1681 TACGGAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
DB 1681 TACGGAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
OY 1741 GGTACTGAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
DB 1741 GGTACTGAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
OY 1801 CTAACTGAAGAGGCTGATGATTTCTAATTAATTAATTAATTAATTAATTAATTAAT 1860
DB 1801 CTAACTGAAGAGGCTGATGATTTCTAATTAATTAATTAATTAATTAATTAATTAAT 1860
OY 1861 ACACCGAAGAGCGAAGCATCTTCCATCTGCTATTCGATTAATTAATTAATTAATTA 1920
DB 1861 ACACCGAAGAGCGAAGCATCTTCCATCTGCTATTCGATTAATTAATTAATTAATTA 1920
OY 1921 CAATCACTTTAGATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
DB 1921 CAATCACTTTAGATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980

1981 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCCTGCAACACA 2040
 1981 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCCTGCAACACA 2040
 1981 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCCTGCAACACA 2040
 2041 TACTTAAATGATCAAAATGATGATTAACATCTGATTAACCAACCAACCAACATTC 2100
 2041 TACTTAAATGATCAAAATGATGATTAACATCTGATTAACCAACCAACCAACATTC 2100
 2101 GGAACAACTGACCAAGTATTTTCAATGATGATGATGATGATGATGATGATGATGAT 2160
 2101 GGAACAACTGACCAAGTATTTTCAATGATGATGATGATGATGATGATGATGATGAT 2160
 2161 ATTAGAACATCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2161 ATTAGAACATCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2221 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 2221 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 2281 TTTCAGTCCACCAAGTATTTTCAATGATGATGATGATGATGATGATGATGATGATGAT 2340
 2281 TTTCAGTCCACCAAGTATTTTCAATGATGATGATGATGATGATGATGATGATGATGAT 2340
 2341 AATTCATCTTATACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 2341 AATTCATCTTATACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 2401 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 2401 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 2461 GAAGGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 2461 GAAGGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 2521 GAGTACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 2521 GAGTACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 2581 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCCTGCAACACA 2640
 2581 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCCTGCAACACA 2640
 2641 GTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
 2641 GTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
 2701 TACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
 2701 TACTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
 2761 TATTAACACTTTGGGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGAT 2820
 2761 TATTAACACTTTGGGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGAT 2820
 2821 GGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880
 2821 GGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880
 2881 AGAAAAATTAACAGCAAGGTTTCAAAACCAATTCAAAAAGTTGATGATGATGATGAT 2940
 2881 AGAAAAATTAACAGCAAGGTTTCAAAACCAATTCAAAAAGTTGATGATGATGATGAT 2940
 2941 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
 2941 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
 3001 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
 3001 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
 3061 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120

3061 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120
 3061 GGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120
 3121 TCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
 3121 TCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
 3181 AAGAAATTA 3189
 3181 AAGAAATTA 3189
 3181 AAGAAATTA 3189

RESULT 2
 AB023292
 LOCUS
 DEFINITION
 AB023292 3189 bp DNA linear BCT 27-MAY-2000
 Mycoplasma gallisepticum gene for 120kDa membrane protein, complete cds.
 ACCESSION
 AB023292
 VERSION
 AB023292.1 GI:4240363
 KEYWORDS
 120kDa membrane protein.
 SOURCE
 Mycoplasma gallisepticum
 ORGANISM
 Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 REFERENCE
 1 (sites)
 Yoshida, S., Fujisawa, A., Tsuzaki, Y. and Saitoh, S.
 Identification and expression of a Mycoplasma gallisepticum surface
 antigen recognized by a monoclonal antibody capable of inhibiting
 both growth and metabolism
 Infect. Immun. 68 (6), 3186-3192 (2000)
 JOURNAL
 MEDLINE
 20278096
 PUBMED
 10816462
 REFERENCE
 2 (bases 1 to 3189)
 Yoshida, S.
 Direct Submission
 Submitted (02-FEB-1999) Shigeto Yoshida, Jichi Medical School,
 Department of Medical Zoology, Yakushiji 3311-1,
 Minamikawachimachi, Tochigi 329-0498, Japan
 (E-mail: shigeto@jichi.ac.jp, Tel: 81-285-58-7339,
 Fax: 81-285-44-6489)

FEATURES
 source
 1..3189
 /organism="Mycoplasma gallisepticum"
 /mol_type="genomic DNA"
 /strain="R"
 /db_xref="taxon:2096"
 1..3189
 /gene="M11"
 1..3189
 /gene="M11"
 /codon_start=-1
 /transl_table=4
 /product="120kDa membrane protein"
 /protein_id="BA7482.1"
 /db_xref="GI:4240363"
 /translation="MNISKSLKSYLIGLAVFAGLGSAPFGKSDKSNDFTOVLNO
 ARTLDANSVRLAGLQNGSLFNTVLDVNDVITANGYIIKIDSTPKLYGLDLSDD
 FAGYVKQIVSYTTSRNFEDROTRAYALVDEANGLIRLNTNSRNNRNS
 KRVIGCVNPAHVIRPTDCTKPFNKQKQGLVNDPILDAPIEDLPDMYNIAYO
 KRLPNDVATVAVPVGVNCGTADNDDGFEDNGOITFTDPIAOTKTTTNDONSTF
 NSGAMGANNRRYDOLNVRHRTSFQLEKRVYPMPTSEENKNTRLATSLPSE
 RYWIIDIPGTPQVTLKEDSVNVSRLSVLSLIGSIIYFGLSELSIMYSFP
 TRLSDELALNQKTDIEASSTIDNGTNTTADSSGSGTGAAGTNTSQTYSN
 PLNTYRSFGIDSKPSASANKIDETNADNVDNVEARIYAEIRGLIONEIPITNAGNFR
 NTIGCVPTSTGSRVYLRASVNDQPTNPOPLVYRGLYQQRCTGTFPGYTKL
 LANSPIYDVLDAARVGETNORFRTSLTFYVMGGLTEEGAKSFNTPIYRAQGDIPES
 RSIFPSGYSDNRYEYIÖSVLGEDGLNNLNVGKASSFLNSNPNGLMLAATVY
 RSQIGLARTSGLPNOQPFSTHVOIVSVSGDQFSSIKNRTIFPGNOLMYPLETNSN
 KSSVYDRLRLDSSNPASSSEPTSLIDNVEIGVILPLDNSFYVNAAGNALFSSN
 PCSPGSTAVNPFNOMLSDIAEFGSAGKATSPDWKGIORPBEYLIONGFTSOVAFN
 VTNOSFLNSLVFTPANAGTNRKYVDPPGNITNOLPLAKQIOYQIDGYYIAKIKNN
 NLVTEYSNPFALPSNVVFPALGSLGIIAIIILDLAIGIPLAKRKLQDGEKTF
 KVDVLTAAVGSVYKRIITQTAIVKRRPALALGKSGDKRPAAPAPAPAPAPAPAPAPAP
 SSPAKPTABKSGAPKPTAPKPAAPKPTAPKE"

Db	961	AAATATTACAAGATTAGCTACTGGAAGTTTGGCAAGCAACGAAAGATATTGAATCTTGGAC	1020
Qy	1021	ATATCCCGGGACATCCCAAGTTACTTTAAAGAAGATTCAGTTAAAGTATTTTCACAGCTA	1080
Db	1021	ATACACGGGGACTCCACAAAGTTACTTTAAAGAAGATTCAGTTAAAGTATTTTCACAGCTA	1080
Qy	1081	TACTTAACAGTAACTTAATCTTTATTCATTCATGTGTAGTATTATTATTTTGGTACC	1140
Db	1081	TACTTAACAGTAACTTAATCTTTATTCATTCATGTGTAGTATTATTATTTTGGTACT	1140
Qy	1141	TCTGAATTTACCATTCATTATGTAGTACTATTCACATCCCAATGATTTTCATCTAACCGCT	1200
Db	1141	TCTGAATTTACCATTCATTATGTAGTACTATTCACATCCCAATGATTTTCATCTAACCGCT	1200
Qy	1201	TTGAATCAAGTTAAACAGATGATTTGAAGCTTCACAGCTGTAATACGGTACACAAACA	1260
Db	1201	TTGAATCAAGTTAAACAGATGATTTGAAGCTTCACAGCTGTAATACGGTACACAAACA	1260
Qy	1261	AAGGAAACACGACACACAGCTGATACATGCTAGCGTTCCACAGGTCGTGAACAGAAAT	1320
Db	1261	AACGGAACACGACACACAGCTGATACATGCTAGCGTTCCACAGGTCGTGAACAGAAAT	1320
Qy	1321	ACTACTAACAATCTCCAAACAGTTTCTAATCCTACTTTAATACCTTATACGATTTTGA	1380
Db	1321	ACTACTAACAATCTCCAAACAGTTTCTAATCCTACTTTAATACCTTATACGATTTTGA	1380
Qy	1381	ATTGATAGTAAACCAACTCTTGCAAAACAAATATGATGAACATTAATTTGGGACATCCTAAC	1440
Db	1381	ATTGATAGTAAACCAACTCTTGCAAAACAAATATGATGAACATTAATTTGGGACATCCTAAT	1440
Qy	1441	GTTATTGAAGCAAGATATATGCTGAATACAGATACAGATTTCAATTCAAAATGAAATTCACATA	1500
Db	1441	GTTATTGAAGCAAGATATATGCTGAATACAGATACAGATTTCAATTCAAAATGAAATTCACATA	1500
Qy	1501	ACTAATGCAAGAACTTTATCCGAAACAAATGGTGGTGGTGGTTTACTTCCACAGGT	1560
Db	1501	ACTAATGCAAGAACTTTATCCGAAACAAATGGTGGTGGTGGTTTACTTCCACAGGT	1560
Qy	1561	TCAAGAGTGGTTTAAAGCTTCTTATTAACGGTGATCAACGTCCAACTGGAAACTTCCAA	1620
Db	1561	TCAAGAGTGGTTTAAAGCTTCTTATTAACGGTGATCAACGTCCAACTGGAAACTTCCAA	1620
Qy	1621	CCTTTCTTATACGTATTTGGTATTTAGGATFACCAACAACTAGAACAGGAACTTCTTGG	1680
Db	1621	CCTTTCTTATACGTATTTGGTATTTAGGATFACCAACAACTAGAACAGGAACTTCTTGA	1680
Qy	1681	TACGGAACATATTAACATTTAAACAACGCCCTTACGACGATTAAGTATGCTGCACAAAGTA	1740
Db	1681	TACGGAACATATTAACATTTAAACAACGCCCTTACGACGATTAAGTATGCTGCACAAAGTA	1740
Qy	1741	GGTACTGAAGAACCAATCATTTTGAAGAAGCTCATTAACATACCTGTTATGGGTGATAT	1800
Db	1741	GGTACTGAAGAACCAATCATTTTGAAGAAGCTCATTAACATACCTGTTATGGGTGATAT	1800
Qy	1801	CTTAAGTGAAGAGTCTAGAAATTTCTTAATACCTCATATATTAAGACCAAGAGTGAC	1860
Db	1801	CTTAAGTGAAGAGTCTAGAAATTTCTTAATACCTCATATATTAAGACCAAGAGTGAC	1860
Qy	1861	ACACCAAGAACCCGAAGCATCTTCCAAATGGCTTCTGATTAATAGTTATGATGATCATT	1920
Db	1861	ACACCAAGAACCCGAAGCATCTTCCAAATGGCTTCTGATTAATAGTTATGATGATCATT	1920
Qy	1921	CAATCAGATTAGGATTTGATGGAATTTAGAAATTAACCTTAATGTTGGGTTTAAAGCTCA	1980
Db	1921	CAATCAGATTAGGATTTGATGGAATTTAGAAATTAACCTTAATGTTGGGTTTAAAGCTCA	1980
Qy	1981	AGCTTCTTAAATCTCAATATAGCAAAATCCAAAGGCTTATGAAGATTTGTGTCGAACACA	2040
Db	1981	AGCTTCTTAAATCTCAATATAGCAAAATCCAAAGGCTTATGAAGATTTGTGTCGAACACA	2040
Qy	2041	TACTTAAAGTACAAGTTGGATTAGCTAGCAACTCTGATTTCCAAACCAACACATTC	2100
Db	2041	TACTTAAAGTACAAGTTGGATTAGCTAGCAACTCTGATTTCCAAACCAACACATTC	2100

OY 2101 GGAGCACTACCAAGTTATTTAGTATCACCCTGGTATCACTTTCATCAATTAGAAAT 2160
 |||||
 Db 2101 GGAGCACTACCAAGTTATTTAGTATCACCCTGGTATCACTTTCATCAATTAGAAAT 2160
 |||||
 OY 2161 ATTAGAACATCTTCCCTGGTACCAAGTTATGCTACTTCTTATTCACAAATGAAT 2220
 |||||
 Db 2161 ATTAGAACATCTTCCCTGGTACCAAGTTATGCTACTTCTTATTCACAAATGAAT 2220
 |||||
 OY 2221 AATGCTAGTTTATACATTATAGATTAGTACTGACTCAAGTACCGTATGCGTCAAGCTTA 2280
 |||||
 Db 2221 AATGCTAGTTTATACATTATAGATTAGTACTGACTCAAGTACCGTATGCGTCAAGCTTA 2280
 |||||
 OY 2281 TTTCAGTCCAAAGTTTAAATGACGTTAAATGAATTTGTGTATCTTATCTTATTAGAC 2340
 |||||
 Db 2281 TTTCAGTCCAAAGTTTAAATGACGTTAAATGAATTTGTGTATCTTATCTTATTAGAC 2340
 |||||
 OY 2341 AATTCTATTATACAGTAAATGCTGCTGTAATTTGTCATTTCTTCATCAAAACCTGGT 2400
 |||||
 Db 2341 AATTCTATTATACAGTAAATGCTGCTGTAATTTGTCATTTCTTCATCAAAACCTGGT 2400
 |||||
 OY 2401 TCTCCGTGATCAATATCTGCTGTAATATCAATTTAATCAGACCTTATCTGATTTT 2460
 |||||
 Db 2401 TCTCCGTGATCAATATCTGCTGTAATATCAATTTAATCAGACCTTATCTGATTTT 2460
 |||||
 OY 2461 GAAGGTTCTGCTGCTAGTATACATCTGATTTCTGGGGAACAATCCAAATCCAGAT 2520
 |||||
 Db 2461 GAAGGTTCTGCTGCTAGTATACATCTGATTTCTGGGGAACAATCCAAATCCAGAT 2520
 |||||
 OY 2521 GAGTACTTATTCAAAAGGCTTCACTAGTCAAGTGGCTAGAAACCTTCTTACAAACCA 2580
 |||||
 Db 2521 GAGTACTTATTCAAAAGGCTTCACTAGTCAAGTGGCTAGAAACCTTCTTACAAACCA 2580
 |||||
 OY 2581 AGCTTCTTAAACAGTTTACTGACTCAGTCCCTGAATGCTGCTGCTAATCACTGCTGA 2640
 |||||
 Db 2581 AGCTTCTTAAACAGTTTACTGACTCAGTCCCTGAATGCTGCTGCTAATCACTGCTGA 2640
 |||||
 OY 2641 GTGGTGTATCTGATGCTAATTTAAACAAACCAAAACCTCTTAAAGTTCCAGATCCAA 2700
 |||||
 Db 2641 GTGGTGTATCTGATGCTAATTTAAACAAACCAAAACCTCTTAAAGTTCCAGATCCAA 2700
 |||||
 OY 2701 TACTTATGATGCTAATTTATGATGCTAAATTTAAAGCAATATTTAGTAACTTCTCT 2760
 |||||
 Db 2701 TACTTATGATGCTAATTTATGATGCTAAATTTAAAGCAATATTTAGTAACTTCTCT 2760
 |||||
 OY 2761 TATTAACACTTTGGCCCTTACCTTCATGGTATGCTTACAGCAATTTGCTATTAATTA 2820
 |||||
 Db 2761 TATTAACACTTTGGCCCTTACCTTCATGGTATGCTTACAGCAATTTGCTATTAATTA 2820
 |||||
 OY 2821 GGTATTTCTGCAATTTGATCATCTTAGATTTAGTATGCTATTCCTTTAAGAGCTCAA 2880
 |||||
 Db 2821 GGTATTTCTGCAATTTGATCATCTTAGATTTAGTATGCTATTCCTTTAAGAGCTCAA 2880
 |||||
 OY 2881 AGAAAATTAACAAGAGGTTCAAACCAATTCAAAAAGTTGATACCTTGACTGCT 2940
 |||||
 Db 2881 AGAAAATTAACAAGAGGTTCAAACCAATTCAAAAAGTTGATACCTTGACTGCT 2940
 |||||
 OY 2941 GCTGTTGTTCACTTACAAAGATTTATACCAACGCTTACCTTAAGAAAAACCT 3000
 |||||
 Db 2941 GCTGTTGTTCACTTACAAAGATTTATACCAACGCTTACCTTAAGAAAAACCT 3000
 |||||
 OY 3001 GCTGTTTATGCTGTTAAATCTGCTGATTAAGAAAAACGCTGCTCTTAACCTGCT 3060
 |||||
 Db 3001 GCTGTTTATGCTGTTAAATCTGCTGATTAAGAAAAACGCTGCTCTTAACCTGCT 3060
 |||||
 OY 3061 GCTCCAGCTAAACATCTGCTGCTGATTAAGAAAAACGCTGCTGCTGCTGCTGCT 3120
 |||||
 Db 3061 GCTCCAGCTAAACATCTGCTGCTGATTAAGAAAAACGCTGCTGCTGCTGCTGCTGCT 3120
 |||||
 OY 3121 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 |||||
 Db 3121 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 |||||

OY 3181 AAGAATTA 3189
 |||||
 Db 3181 AAGAATTA 3189
 |||||
 RESULT 3
 AEO16967
 LOCUS
 DEFINITION
 Mycoplasma gallisepticum strain R section 1 of 4 of the complete genome
 ACCESSION
 AEO16967 AEO15450
 VERSION
 AEO16967.1 GI:31541048
 KEYWORDS
 SOURCE
 ORGANISM
 Mycoplasma gallisepticum R
 Bacteria; Filicutes; Mollicutes; Mycoplasmatellaceae; Mycoplasma.
 REFERENCE
 1 (bases 1 to 301042)
 Geary,S.J., Papazisi,L., Kutish,G., Gorton,T.S., Mahairas,G., Swartzell,S., Maden,A., Nguyen,D.K., Markham,P., Browning,G., Kamal,M. and Liao,X.
 The complete genome sequence of the avian pathogen Mycoplasma gallisepticum strain R
 Microbiology (2003) In press
 2 (bases 1 to 301042)
 Geary,S.J., Papazisi,L., Kutish,G., Mahairas,G., Swartzell,S., Maden,A., Nguyen,D.K., Gorton,T.S., Markham,P., Browning,G., Mustafa,K. and Liao,X.
 Direct Submission
 TITLE
 Submitted (17-OCT-2002) Department of Pathobiology and Veterinary Sciences, and Center of Excellence for Vaccine Research, The University of Connecticut, 61 North Eagleville Road U-89, Storrs, CT 06269-3089, USA
 JOURNAL
 FEATURES
 source
 1. 301042
 location/Qualifiers
 /organism="Mycoplasma gallisepticum R"
 /mol_type="genomic DNA"
 /strain="R"
 /db_xref="taxon:233150"
 1. 1889
 rep_origin
 /note="putative oric"
 complement(24. .32)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma pulmonis consensus titacmama"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 complement(36. .44)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma pulmonis consensus titacmama"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 114. .122
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 209. .217
 /note="oric imperfect repeat"
 complement(254. .262)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma mycoides consensus YtWMHAMR"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 298. .306
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma mycoides consensus YtWMHAMR"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 complement(401. .450)
 /note="oric imperfect repeat"
 repeat_region

QY	121	TTAGTTAATCAAGCAGAACCGCTAGATGCTAAATTCCTGTTAGACTTGCAGGCTCTTGGACAA	180
Db	226727	TTAGTTAATCAAGCAGAACCGCTAGATGCTAAATTCCTGTTAGACTTGCAGGCTCTTGGACAA	226786
QY	181	AATGGTCGTGTTCAATACAGTCTTAGAGTGTGATGATGAATCTTATACACACAGCT	240
Db	226787	AATGGTCGTGTTCAATACAGTCTTAGAGTGTGATGATGAATCTTATACACACAGCT	226846
QY	241	AATGAGCAATATCAAAATAGATAGTATTTACTAAACAATATATGTTAGATGCTAAAGT	300
Db	226847	AATGAGCAATATCAAAATAGATAGTATTTACTAAACAATATATGTTAGATGCTAAAGT	226906
QY	301	GATGATTCGTGCTGATACAAAGTAACAAATAGTTTCAGATTACACAACATGCAGAAAT	360
Db	226907	GATGATTCGTGCTGATACAAAGTAACAAATAGTTTCAGATTACACAACATGCAGAAAT	226966
QY	361	AGATTTCATCAAGCACAACAGACATATTAATGCTCTGTTGCTTAATGATGACGCTAAC	420
Db	226967	AGATTTCATCAAGCACAACAGACATATTAATGCTCTGTTGCTTAATGATGACGCTAAC	227026
QY	421	GTCATATTAACAAAAGATTAATACATACATCAATTAATTTGGTAATAGAAACAACAATCT	480
Db	227027	GTCATATTAACAAAAGATTAATACATACATCAATTAATTTGGTAATAGAAACAACAATCT	227086
QY	481	AAGTTTGTAAATTTGGTGTGTGATTAATCCAGCTCAGTAATTAAGATTTACTGATGAGG	540
Db	227087	AAGTTTGTAAATTTGGTGTGTGATTAATCCAGCTCAGTAATTAAGATTTACTGATGAGG	227146
QY	541	ACTAAATTAATTTTACCAACCACTCAAGGTGAATTTGTTAATGACTTCAATTTAGAT	600
Db	227147	ACTAAATTAATTTTACCAACCACTCAAGGTGAATTTGTTAATGACTTCAATTTAGAT	227206
QY	601	GGCGCAATCTTACCTAAAGATTTACACCAAGTGTGTAATCTTATACCTTAACAAAGAAG	660
Db	227207	GGCGCAATCTTACCTAAAGATTTACACCAAGTGTGTAATCTTATACCTTAACAAAGAAG	227266
QY	661	ATCTTACCAATGAGCTCAGACACTGCTCAGTTGTTCTTGCCAGTAGTAGTATAGTGA	720
Db	227267	ATCTTACCAATGAGCTCAGACACTGCTCAGTTGTTCTTGCCAGTAGTAGTATAGTGA	227326
QY	721	ACAAATGCTGATGATGGATGTTTGATTTGTGGGAATGCTCAATAATACATATACAGATCT	780
Db	227327	ACAAATGCTGATGATGGATGTTTGATTTGTGGGAATGCTCAATAATACATATACAGATCT	227386
QY	781	ATFGTCAAACTAAACACACTACGTATATACAAATCCCTCAACTTTTAACTACAGAGCA	840
Db	227387	ATFGTCAAACTAAACACACTACGTATATACAAATCCCTCAACTTTTAACTACAGAGCA	227446
QY	841	ATGCGTGGTGCACAAATAGATACGATTTCTCAATGGAATGTACAGCATAGAAATTAACA	900
Db	227447	ATGCGTGGTGCACAAATAGATACGATTTCTCAATGGAATGTACAGCATAGAAATTAACA	227506
QY	901	TCCTTCCAAATAGATGAACAAATTTGTTTATCCAGATGACCTGGTCTGAGAGAAATAAA	960
Db	227507	TCCTTCCAAATAGATGAACAAATTTGTTTATCCAGATGACCTGGTCTGAGAGAAATAAA	227566
QY	961	AATATTAACAAGTTGCTACTAGGGAAGTTGCCAACCAACGAAGAATATGGATCTCTGAC	1020
Db	227567	AATATTAACAAGTTGCTACTAGGGAAGTTGCCAACCAACGAAGAATATGGATCTCTGAC	227626
QY	1021	ATACCGGAGACCTCCACAGTACTTAAATAAGAAATTCAGTTAACGTATTTTCAAGACAT	1080
Db	227627	ATACCGGAGACCTCCACAGTACTTAAATAAGAAATTCAGTTAACGTATTTTCAAGACAT	227686
QY	1081	TACTTAAACCTCAGTTAATTCCTTATCATTCATTTGGTATGATATTAATTTTGGTACC	1140
Db	227687	TACTTAAACCTCAGTTAATTCCTTATCATTCATTTGGTATGATATTAATTTTGGTACC	227746
QY	1141	TCTGAATACCATCATATATGATACATTCATTTCCCAACATAGATATATCTATCTAACCGCT	1200
Db	227747	TCTGAATACCATCATATATGATACATTCATTTCCCAACATAGATATATCTATCTAACCGCT	227806

QY	1201	TTGATCAAGTTAAAAAGATATATTTGAAGCTTCACAGCAGTATACGGTACAAACAA	1260
Db	227807	TTGAATCAAGTTAAAAAGATATATTTGAAGCTTCACAGCAGTATACGGTACAAACAA	227866
QY	1261	AACGAAACAGCAACAGCTGATTCATCTAGCTGGTTCACAGGCTCTGGAACAGGAAT	1320
Db	227867	AACGAAACAGCAACAGCTGATTCATCTAGCTGGTTCACAGGCTCTGGAACAGGAAT	227926
QY	1321	ACTACTAACACTTCTCAAAACAGTTTCTAATCCTACTTTTAAATACATTCATGCTTGTGGA	1380
Db	227927	ACTACTAACACTTCTCAAAACAGTTTCTAATCCTACTTTTAAATACATTCATGCTTGTGGA	227986
QY	1381	ATTGATAGTAACCAACTTCTGCAACAAATATAGATGAATGGCAGATCTCAAC	1440
Db	227987	ATTGATAGTAACCAACTTCTGCAACAAATATAGATGAATGGCAGATCTCAAT	228046
QY	1441	GTTATTTGAAGCAAGAATATATGCTGAATACATTAAGTATATCAAAATGCAATTCGAAAT	1500
Db	228047	GTTATTTGAAGCAAGAATATATGCTGAATACATTAAGTATATCAAAATGCAATTCGAAAT	228106
QY	1501	ACTAATGCAAGAAACTTTATCCGAAACACAATGGGCGTTGGTTTACTTCAACAGGT	1560
Db	228107	ACTAATGCAAGAAACTTTATCCGAAACACAATGGGCGTTGGTTTACTTCAACAGGT	228166
QY	1561	TCACAGATGATTTTAAAGCCTTCTATTAACGGTGATCAACGTCCAACTGSAACCTTCCAA	1620
Db	228167	TCACAGATGATTTTAAAGCCTTCTATTAACGGTGATCAACGTCCAACTGSAACCTTCCAA	228226
QY	1621	CCTTCTCTATACGTATTTGGTTATTTAGATATCCAAACAACTATGAAACAGGAACTTTCTGG	1680
Db	228227	CCTTCTCTATACGTATTTGGTTATTTAGATATCCAAACAACTATGAAACAGGAACTTTCTGG	228286
QY	1681	TACGGAACATATACGCTTTTAAACAAACAGCCCTTACAGCGTATTTAGATTTCCCAAGATA	1740
Db	228287	TACGGAACATATATACCTATTTAAACAAACAGCCCTTACAGCGTATTTAGATTTCCCAAGATA	228346
QY	1741	GGTACTGAAACCAATCAATTTTGAAGAACCTTATTAACATCCCTGTTATGGGTGATAT	1800
Db	228347	GGTACTGAAACCAATCAATTTTGAAGAACCTTATTAACATCCCTGTTATGGGTGATAT	228406
QY	1801	CTAATCTGAAGAAGTGTCTGAAGACTTTCCTATATCTCCATATATTAAGACACAAAGGTGAC	1860
Db	228407	CTAATCTGAAGAAGTGTCTGAAGACTTTCCTATATCTCCATATATTAAGACACAAAGGTGAC	228466
QY	1861	ACACCAGAAAGCCGAAGCATCTTCCAAATCGCTATTCGATTAATCTTATGAGTACAT	1920
Db	228467	ACACCAGAAAGCCGAAGCATCTTCCAAATCGCTATTCGATTAATCTTATGAGTACAT	228526
QY	1921	CAATCAGTTTATAGATTTTGATGGAATTTAGAAATTAACTTAAATGTTGGGTTTAAAGCATCA	1980
Db	228527	CAATCAGTTTATAGATTTTGATGGAATTTAGAAATTAACTTAAATGTTGGGTTTAAAGCATCA	228586
QY	1981	AGCTTCTTAACTCAAAATAGACCAAAATCCAAAGGCTTGAAGAAATGTTGCTGCAACAA	2040
Db	228587	AGCTTCTTAACTCAAAATAGACCAAAATCCAAAGGCTTGAAGAAATGTTGCTGCAACAA	228646
QY	2041	TACTTAAGATTCACAAATTTGGATTAGCTAGACATCTGGATTTACCAACCAACCAACATTC	2100
Db	228647	TACTTAAGATTCACAAATTTGGATTAGCTAGACATCTGGATTTACCAACCAACCAACATTC	228706
QY	2101	GGACAAACTCACCAAGTTATTTTCAGATATCACCTGGTATCAGTTCTCATCAATTAAGAAAT	2160
Db	228707	GGACAAACTCACCAAGTTATTTTCAGATATCACCTGGTATCAGTTCTCATCAATTAAGAAAT	228766
QY	2161	ATTAGAACATCTTCCCTGTGTAAACCGATTATGTAATCTTCTTATTTCAACAAATGAATTAAT	2220
Db	228767	ATTAGAACATCTTCCCTGTGTAAACCGATTATGTAATCTTCTTATTTCAACAAATGAATTAAT	228826
QY	2221	AAATCTAGTGTATATCATTAAGATATAGCTACAGTCAAGTAACCTATAGCGTCAACCTCA	2280
Db	228827	AAATCTAGTGTATATCATTAAGATATAGCTACAGTCAAGTAACCTATAGCGTCAACCTCA	228886
QY	2281	TTTCAGTCCCAACAGTTTAAATGACGTTTAAATGAATTTGTTATTAATCTTATCTTATTAGAC	2340

[illegible]

		/translation-"N1SKRLKSYTLIGLAVGALGSASFGRKSGKNDNTOLVNO ARTLDANSVRLAGLNGSLPTVLDVDDNFTNANGTIIKIDSTPKLYGLDSD FAGYKVOJVSYYTTSRNRDROTRAYVALLVNDANGLIKDSTNRIGRRNNS KEVIGGVNDPAHVIRTDGTFKFNFKQKOGELVNDPILDAPIPLDLPDNLXIQ KRLPNDVNTAVVPMFVGRVSGTNADGDFDNGOITNDPDKOCTTNDONPSTE NSGAMGANNRVSQNLNVRHRIKTSQLEKFPVPEMTSEBKNTLRATLGSLENE RWIILDIPGTPVTLKEDSVNFSRLNSVNSLSFGISYIFGSESLPMYVFP TRLSDLTALNOVKTDIEASTDNGTTTDTTDSSTSGAGGNTNDSQVSN PILNMYRSFGIDSKPSAKIDETNADNVI EARLYEARLIGIONEIDITNANIR NLTGCGYPTSGSRVLRASVNGDOPRTGNPQPLVREGLYQOIRTEPTWGTIKL LANSPIVDVDAKRVETINQFRISLTIVMGGLTEBEARSESNPIYTRAGDTEF RSIFOSGSDNTY EYIOVLGFDGIRNNLVNGKASSFNNSNNGLEMAAATYL RSOIGLARTSGLPNOQPFCTHOVISFGDQFSIKNRTIEFGNOLWFLENTYL KSSVYTLRLADSNPDASSPSPTSLIDVNEIGLILPLDNSPYTAAANGVALFSSN PSPGSGYTAVENTENOMLSDAFEGSGAKYTSDEMGTIOPKPDXYLIONEPTSOVARN VITNOSFLNSVDPTPANACTNRYVYVDPGNGELNOLPLKVOIYIDGXYDKKLN NLVTFYSNNFALPBNVPTAIGSTIGIILAILLILDLAIGLIPRAKRIQDDKFTFE KAVDTLTAAVGSYKRIIYQTAHVKKRPALGKSGDKPAAAPAPAPRSABKA SSPAKPTAKSGAPTKPAPKAPAPKPAKE"		7128 . 8126 /gene-"crmb" /note-"Crimb; similar to N-terminus of Mycoplasma gallisepticum Gapa" /codon_start=1 /transl_table=4 /product="putative cytochrome related protein B" /db_xref="GI:6694737" /translation-"MLKAHRSLLIFLILIAISLIEIGLEFKKPNINOVLKNO SSSLITNOPSDIIRKYSLSDIRMNNANFSSPTKDGILVITTEGRERISFYL KMELANRNINLATVDEONLVSDNRITKRPDAVEINNKSYYAAVIAIOSDEFTI XYLIASDOLLINDLKDLISFOGLATDLTSGVLAIVKVENPDISDPGCIIDYAMIS PHDIDONTYPAKMTNTLSIXGDNBSFRKRDHPSWFNDNSFISLPMLOQITNLANIE YDNVFLFSGNGWVAIVFVSKOPNTFLGLESGLIKIIVLDLSOVLMLTY"	
BASE COUNT	2916 a	1463 c	1426 g	2545 t	4 others
ORIGIN					
Query Match	98.6%	Score 3143.4	DB 1	Length 8354	
Best Local Similarity	99.1%	Pred. No. 0			
Matches 3159	Conservative	2	Mismatches	28	Indels 0; Gaps 0;
OY	1	ATGAATATTTCTAAAAACCTTAAAGTTATACATGTAGTAGTGATTAGCTGTATTGGA	60		
DB	3778	ATGAATATTTCTAAAAACCTTAAAGTTATACATGTAGTAGTGATTAGCTGTATTGGA	3837		
OY	61	GCTCTTGCTTCGCAAGCTTTGGCTTTAAGCAATCAGATPAGAGTAAACGATTAACGCGAA	120		
DB	3838	GCTCTTGCTTCGCAAGCTTTGGCTTTAAGCAATCAGATPAGAGTAAACGATTAACGCGAA	3897		
OY	121	TTAGTTAATCAAGCAAGCGTAGATGCTAATTCGTTAGACTTCGAGCTCTTGACAA	180		
DB	3898	TTAGTTAATCAAGCAAGCGTAGATGCTAATTCGTTAGACTTCGAGCTCTTGACAA	3957		
OY	181	AATGCTCGTTGTTCAATACAGTCTTGAAGATGTTGANGATTAACCTTATTAACGACGT	240		
DB	3958	AATGCTCGTTGTTCAATACAGTCTTGAAGATGTTGANGATTAACCTTATTAACGACGT	4017		
OY	241	AATGSAACAATTTCAAAATTTAGATTCTTAAACCATTAATGTTTAACTTAACTTAACT	300		
DB	4018	AATGSAACAATTTCAAAATTTAGATTCTTAAACCATTAATGTTTAACTTAACTTAACT	4077		
OY	301	GATGATTGCTGATGATCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA	360		
DB	4078	GATGATTGCTGATGATCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA	4137		
OY	361	AGATTGATCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA	420		
DB	4138	AGATTGATCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA	4197		
OY	421	GTTCAATTTAAAGCAATTAATACTCAAAATAGAAATGTAATGTAAGTAAGTAAGTA	480		
DB	4198	GTTCAATTTAAAGCAATTAATACTCAAAATAGAAATGTAATGTAAGTAAGTAAGTA	4257		
OY	481	AATGTTGATTTGCTGCTGTTGATTAATCCAGCTCAGTAAATGATTTACTGATGATGG	540		
DB	4258	AATGTTGATTTGCTGCTGTTGATTAATCCAGCTCAGTAAATGATTTACTGATGATGG	4317		
OY	541	ACAAATTTTAAATTTTCAAAACCAACCTCAAGGTAATGTTAATGACCTTATTAATGAT	600		
DB	4318	ACAAATTTTAAATTTTCAAAACCAACCTCAAGGTAATGTTAATGACCTTATTAATGAT	4377		
OY	601	GGCCCAATCTTACCTTAAAGATTTTACACCCAGATTGGTATPACTTATPACTTAAAGAA	660		
DB	4378	GGCCCAATCTTACCTTAAAGATTTTACACCCAGATTGGTATPACTTATPACTTAAAGAA	4437		
OY	661	ATCTTACCAATTAACCTCAACAGTACGAGTTGCTTCCCTGGCCATAGGTAGATTAGCA	720		
DB	4438	ATCTTACCAATTAACCTCAACAGTACGAGTTGCTTCCCTGGCCATAGGTAGATTAGCA	4497		
OY	721	ACAAATGCTGATGATGGATGTTGATTTGGGAATGTCAAATTAACCTAATACAGATCCT	780		
DB	4498	ACAAATGCTGATGATGGATGTTGATTTGGGAATGTCAAATTAACCTAATACAGATCCT	4557		
OY	781	ATTGCTCAAACTAAACCACTACTGATATCAAAATCTTCAACTTTTAAATTCAGAGCA	840		
DB	4558	ATTGCTCAAACTAAACCACTACTGATATCAAAATCTTCAACTTTTAAATTCAGAGCA	4617		
OY	841	ATGCCGTCGCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	900		
DB	4618	ATGCCGTCGCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	4677		
OY	901	TCTTTCCAAATTAAGTAAAAAATTTGTTTATCCGAATGAGCTGTTCTGAAGAAATTA	960		
DB	4678	TCTTTCCAAATTAAGTAAAAAATTTGTTTATCCGAATGAGCTGTTCTGAAGAAATTA	4737		
OY	961	AATATTACAAGATTAGCTACTGCAAGTTTCCCAAGCAAGCAAGCAAGCAAGCAAGCA	1020		
DB	4738	AATATTACAAGATTAGCTACTGCAAGTTTCCCAAGCAAGCAAGCAAGCAAGCAAGCA	4797		
OY	1021	ATACCGGGGACCTCACAGTACTTTAAAGAAAGATTCACTTAACGATTTTCAAGCTA	1080		
DB	4798	ATACCGGGGACCTCACAGTACTTTAAAGAAAGATTCACTTAACGATTTTCAAGCTA	4857		
OY	1081	TACTTAACTCACTTAATCTTATCATTCATTGTTGGTATPAGATTTATTTTGGTACC	1140		
DB	4858	TACTTAACTCACTTAATCTTATCATTCATTGTTGGTATPAGATTTATTTTGGTACC	4917		
OY	1141	TCTGAATTTACATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATG	1200		
DB	4918	TCTGAATTTACATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATG	4977		
OY	1201	TTGAATCAAGTTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATG	1260		
DB	4978	TTGAATCAAGTTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATG	5037		
OY	1261	AACGGAACAAGCAACAGCTGATACATCTAGTGGTTCAACAGCTGCTGGAACAGAAAT	1320		
DB	5038	AACGGAACAAGCAACAGCTGATACATCTAGTGGTTCAACAGCTGCTGGAACAGAAAT	5097		
OY	1321	ACTACTACACTCTTCAACAGTTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTT	1380		
DB	5098	ACTACTACACTCTTCAACAGTTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTT	5157		
OY	1381	ATTGATAGTTAAACCACTTCTGCAACAAATAGATGAAATTAATGGGAGATTCCTAAC	1440		
DB	5158	ATTGATAGTTAAACCACTTCTGCAACAAATAGATGAAATTAATGGGAGATTCCTAAC	5217		
OY	1441	GTTATTGAGCAAGATATATGCTGAATACAGATTAGATTGTTCAAAATGAAATTCATA	1500		
DB	5218	GTTATTGAGCAAGATATATGCTGAATACAGATTAGATTGTTCAAAATGAAATTCATA	5277		
OY	1501	ACTAATGCAAGAACTTATCCGAAACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1560		
DB	5278	ACTAATGCAAGAACTTATCCGAAACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5337		

QY 1561 TCAGAGTAGTGTAAAGAGCTTCTTAATACGGTATGATCAACGTCACAACTGGAACCTCCAA 1620
 |||||
 Db 5338 TCAGAGTAGTGTAAAGAGCTTCTTAATACGGTATGATCAACGTCACAACTGGAACCTCCAA 5397
 |||||
 QY 1621 CCTTCTTAATAGTATTTGGTATTTAGATACCAACAACTAGAACGAGACTTCTGG 1680
 |||||
 Db 5398 CCTTCTTAATAGTATTTGGTATTTAGATACCAACAACTAGAACGAGACTTCTGG 5457
 |||||
 QY 1681 TAGGAACTATTAAGCTTTTAAACACAGCCCTTACAGAGTATAGATTTCCCAAGATA 1740
 |||||
 Db 5458 TAGGAACTATTAAGCTTTTAAACACAGCCCTTACAGAGTATAGATTTCCCAAGATA 5517
 |||||
 QY 1741 GGTACGAAACCAATCAATTTAGAGAACTTCATTACATACCCCTGTTAGGGGTGAT 1800
 |||||
 Db 5518 GGTACGAAACCAATCAATTTAGAGAACTTCATTACATACCCCTGTTAGGGGTGAT 5577
 |||||
 QY 1801 CTAACGTGAAGAGGTGCTAGAGATTCTCTAATACCTCCATATATATAGAGCAAGGTGAC 1860
 |||||
 Db 5578 CTAACGTGAAGAGGTGCTAGAGATTCTCTAATACCTCCATATATATAGAGCAAGGTGAC 5637
 |||||
 QY 1861 ACACGAGAAAGCGAGACATCTCCATCTGGCTATTTGATATATCTTATAGTACAT 1920
 |||||
 Db 5638 ACACGAGAAAGCGAGACATCTCCATCTGGCTATTTGATATATCTTATAGTACAT 5697
 |||||
 QY 1921 CAATCAGTTTTAGGATTTGATGAAATTAAGAAATTAATGTTGGGTTAAACATCA 1980
 |||||
 Db 5698 CAATCAGTTTTAGGATTTGATGAAATTAAGAAATTAATGTTGGGTTAAACATCA 5757
 |||||
 QY 1981 AGCTCTTAACACTCAATAGACAAATCAACGCGTGTAGAAATGATTTGCTGCACACCA 2040
 |||||
 Db 5758 AGCTCTTAACACTCAATAGACAAATCAACGCGTGTAGAAATGATTTGCTGCACACCA 2100
 |||||
 QY 2041 TACTTAAGATCAAAATTTGATGATGCTAGAAACATCTGATACCAACCAACCACTTC 2100
 |||||
 Db 5818 TACTTAAGATCAAAATTTGATGATGCTAGAAACATCTGATACCAACCAACCACTTC 5877
 |||||
 QY 2101 GGAAACAATCACCAGATTATTCAGATACACCTGGTGTAGTCACTTCTCATTAATTAAGAT 2160
 |||||
 Db 5878 GGAAACAATCACCAGATTATTCAGATACACCTGGTGTAGTCACTTCTCATTAATTAAGAT 5937
 |||||
 QY 2161 ATTAGAACATCTTCCCTGTATACAGATGATGATCTTCTTATTCACAAATGAATAAT 2220
 |||||
 Db 5938 ATTAGAACATCTTCCCTGTATACAGATGATGATCTTCTTATTCACAAATGAATAAT 5997
 |||||
 QY 2221 AAATCTAGTGTATACATTAAGATTAGTGAATGCTGATCAAGTACCCCGATGGTCAAGCTCA 2280
 |||||
 Db 5998 AAATCTAGTGTATACATTAAGATTAGTGAATGCTGATCAAGTACCCCGATGGTCAAGCTCA 6057
 |||||
 QY 2281 TTCAGTCCAAAGTTAAATTTGAGTAAATGCTGATCTTACCTTATTTAGAC 2340
 |||||
 Db 6058 TTCAGTCCAAAGTTAAATTTGAGTAAATGCTGATCTTACCTTATTTAGAC 6117
 |||||
 QY 2341 AATTCATCTATACAGTAATGCTGCTGATATGTTGCAATGTTCTCATCAAAACCTGGT 2400
 |||||
 Db 6118 AATTCATCTATACAGTAATGCTGCTGATATGTTGCAATGTTCTCATCAAAACCTGGT 6177
 |||||
 QY 2401 TCTCTGATCATATCTGCTGATTAATACATTAATCAAGAACTTATCTGATTTGCTTT 2460
 |||||
 Db 6178 TCTCTGATCATATCTGCTGATTAATACATTAATCAAGAACTTATCTGATTTGCTTT 6237
 |||||
 QY 2461 GAAGGTCTGCTAGTATATACATCTGATTTCTGGGAAACATCAATTCAAACCGAT 2520
 |||||
 Db 6238 GAAGGTCTGCTAGTATATACATCTGATTTCTGGGAAACATCAATTCAAACCGAT 6297
 |||||
 QY 2521 GAGTACTTAATCAAAATGGGTCTAGTCAAGTGGTGAACAACTTGTACAAACCA 2580
 |||||
 Db 6298 GAGTACTTAATCAAAATGGGTCTAGTCAAGTGGTGAACAACTTGTACAAACCA 6357
 |||||
 QY 2581 AGCTCTTAACAGTTTACTGATCTCTGCTGATATGCTGAGTAACTACTACCTGTA 2640
 |||||
 Db 6358 AGCTCTTAACAGTTTACTGATCTCTGCTGATATGCTGAGTAACTACTACCTGTA 6417
 |||||

QY 2641 GTGGTGAATCTGATGATATTTAAACAAACCAACCTACTCTAAAGTTCAGATCAA 2700
 |||||
 Db 6418 GTGGTGAATCTGATGATATTTAAACAAACCAACCTACTCTAAAGTTCAGATCAA 6477
 |||||
 QY 2701 TACTTGAATGATGATATTTATGATGCTAAATTTAAAGCAATATTTAGTACATCTCT 2760
 |||||
 Db 6478 TACTTGAATGATGATATTTATGATGCTAAATTTAAAGCAATATTTAGTACATCTCT 6537
 |||||
 QY 2761 TATACAACTTTGGGCTTTACTTCAATGGGATGAGGCTACAGCAATTTAGTACATTA 2820
 |||||
 Db 6538 TATACAACTTTGGGCTTTACTTCAATGGGATGAGGCTACAGCAATTTAGTACATTA 6597
 |||||
 QY 2821 GGTATCTTGGCAATTTATGATCATCTTATAGATTTAGTATGCTATTCCTTTAGAGCTCAA 2880
 |||||
 Db 6598 GGTATCTTGGCAATTTATGATCATCTTATAGATTTAGTATGCTATTCCTTTAGAGCTCAA 6657
 |||||
 QY 2881 AGAAATTTACAAAGCAAGGTTCAAAACCAACATTCAAAAGTTGATACCTTGACTGCT 2940
 |||||
 Db 6658 AGAAATTTACAAAGCAAGGTTCAAAACCAACATTCAAAAGTTGATACCTTGACTGCT 6717
 |||||
 QY 2941 GCTGTGGTTCAGTTTACAAAGATTTATACCAAACTGCTTACGTTAAAGAAACCT 3000
 |||||
 Db 6718 GCTGTGGTTCAGTTTACAAAGATTTATACCAAACTGCTTACGTTAAAGAAACCT 6777
 |||||
 QY 3001 GCTGCTTTAGTGTGCTGATTAATCTGTGATTAAGAAACCTGCTGCTTAAACCTGCT 3060
 |||||
 Db 6778 GCTGCTTTAGTGTGCTGATTAATCTGTGATTAAGAAACCTGCTGCTTAAACCTGCT 6837
 |||||
 QY 3061 GCTCCAGCTTAACCAATCTGACCAAAAGCTAGCTCACCAGCTTAACCAACCTGGGCTTAA 3120
 |||||
 Db 6838 GCTCCAGCTTAACCAATCTGACCAAAAGCTAGCTCACCAGCTTAACCAACCTGGGCTTAA 6897
 |||||
 QY 3121 TCTGTGGCTTACAAACCAACTGCTGTATAGCCAGTGTCTCCAAACCAACCTGCTCC 3180
 |||||
 Db 6898 TCTGTGGCTTACAAACCAACTGCTGTATAGCCAGTGTCTCCAAACCAACCTGCTCC 6957
 |||||
 QY 3181 AAAGATTA 3189
 |||||
 Db 6958 AAAGATTA 6966
 |||||

RESULT 5
 AX113685 10651 bp DNA linear PAT 01-MAY-2001
 LOCUS Sequence 1 from Patent WO0127254.
 DEFINITION AX113685
 ACCESSION AX113685
 VERSION AX113685.1 GI:13939855
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE
 1
 AUTHORS Schwarzmann, F.
 TITLE Gene transfer vectors for treating autoimmune diseases and diseases with immunopathogenesis by therapy
 JOURNAL Patent: WO 0127254-A 1 19-APR-2001;
 Schwarzmam, Fritz (DE)
 FEATURES
 1 location/Qualifiers
 source
 1.10651
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="plasmid"

BASE COUNT 2782 a 2618 c 2490 g 2759 t 2 others
 ORIGIN
 Query Match 98.6%; Score 3143.4; DB 6; Length 10651;
 Best Local Similarity 99.18; Pred. No. 0;
 Matches 3159; Conservative 2; Mismatches 28; Indels 0; Gaps 0;

QY 1 ATGAATATTTCTAATAAACTTAAAGTATATACCTTATAGTATGAGTGGATTTAGCTGATTTTGA 60
 |||||
 Db 6 ATGAATATTTCTAATAAACTTAAAGTATATACCTTATAGTATGAGTGGATTTAGCTGATTTTGA 65
 |||||

OY	61	GCTCTGGTTCGCAAGCTTTGGCTTTAAGCAATACAGTAAGAGTAACGATTAACACGCA	120
Dp	66	GCTCTGGTTCGCAAGCTTTGGCTTTAAGCAATACAGTAAGAGTAACGATTAACACGCA	125
OY	121	TTACTTATCAAGCAAGACGCTGAGTGCCTAACTCTGTATACCTTGCAAGCTTGACAA	180
Dp	126	TTACTTATCAAGCAAGACGCTGAGTGCCTAACTCTGTATACCTTGCAAGCTTGACAA	185
OY	181	AATGGTTTGGTTTCAATACAGTTCTTAAAGATGTTGATGATTAACCTTATATACAGCACT	240
Dp	186	AATGGTTTGGTTTCAATACAGTTCTTAAAGATGTTGATGATTAACCTTATATACAGCACT	245
OY	241	AATGGAACAATTAATCAAAATAGATAGTTTCTAAACCATATATATAGTTTGATCTAAT	300
Dp	246	AATGGAACAATTAATCAAAATAGATAGTTTCTAAACCATATATATAGTTTGATCTAAT	305
OY	301	GATGATTTGGTGGATACCAAAAGTAATAAACCAATAGTTTCAGATTACACAACAGAGAAAT	360
Dp	306	GATGATTTGGTGGATACCAAAAGTAATAAACCAATAGTTTCAGATTACACAACAGAGAAAT	365
OY	361	AGATTGATCAACAACAACAAGACATATATATCTCTGTGTGTTAATGATGAAGCTAAC	420
Dp	366	AGATTGATCAACAACAACAAGACATATATATCTCTGTGTGTTAATGATGAAGCTAAC	425
OY	421	GTTCAATTTAAAAAATTAATTAATCTAATCAATAGAAATTTGTAATAGAAACAACAATCT	480
Dp	426	GTTCAATTTAAAAAATTAATTAATCTAATCAATAGAAATTTGTAATAGAAACAACAATCT	485
OY	481	AAGTTTGTAAATTTGGTGTGTGATATACACGCTACGCTAATTAAGATTTACTGATGATGG	540
Dp	486	AAGTTTGTAAATTTGGTGTGTGATATACACGCTACGCTAATTAAGATTTACTGATGATGG	545
OY	541	ACTAAATTTAATTTTACAAACCAACATCAAGTGAATTTGTAATAGATTCATTTAGAT	600
Dp	546	ACTAAATTTAATTTTACAAACCAACATCAAGTGAATTTGTAATAGATTCATTTAGAT	605
OY	601	GGGCCAATCTTACCTAAAGATTACACCCAGATTGTAATTAATTAATCAATTCAAAGAAAG	660
Dp	606	GGGCCAATCTTACCTAAAGATTACACCCAGATTGTAATTAATTAATCAATTCAAAGAAAG	665
OY	661	ATCTTACCAATGAGCGTCAACACGCAAGTCTTCTGGCCAGTAGAGTAGTAGTGA	720
Dp	666	ATCTTACCAATGAGCGTCAACACGCAAGTCTTCTGGCCAGTAGAGTAGTAGTGA	725
OY	721	ACAATAGCTGATGATGGGATGTTTGATTTGGGATGGTCAAAATTAATCAATACAGATCT	780
Dp	726	ACAATAGCTGATGATGGGATGTTTGATTTGGGATGGTCAAAATTAATCAATACAGATCT	785
OY	781	ATTGCTCAAACTAAAACACACTACTGATATATAAATCTTCACTTTTAAATTCAGAGCA	840
Dp	786	ATTGCTCAAACTAAAACACACTACTGATATATAAATCTTCACTTTTAAATTCAGAGCA	845
OY	841	ATGCTGATGCAAAACAATAGATAGATTTCTCAATGGATGTCGAAGCATAGAAATTAACA	900
Dp	846	ATGCTGATGCAAAACAATAGATAGATTTCTCAATGGATGTCGAAGCATAGAAATTAACA	905
OY	901	TCTTTCCAAATTAAGATGAAAAATTTGTTTATCCAGATAGGACTGGTTCTGAAGAGAAATAA	960
Dp	906	TCTTTCCAAATTAAGATGAAAAATTTGTTTATCCAGATAGGACTGGTTCTGAAGAGAAATAA	965
OY	961	AATATTTACAAATTAATAGTACTGGAAGTTTGGCAAGCAAGAAAGATATGGATCTTAC	1020
Dp	966	AATATTTACAAATTAATAGTACTGGAAGTTTGGCAAGCAAGAAAGATATGGATCTTAC	1025
OY	1021	ATACCGGAGATCCACAAGTACTTTTAAAAAAGATTAAGTTAAGTTTTCACAAGACTA	1080
Dp	1026	ATACCGGAGATCCACAAGTACTTTTAAAAAAGATTAAGTTAAGTTTTCACAAGACTA	1085
OY	1081	TACTTAAACTCAGTTAATCTTTATATCATTCATTTGGTGATAGTATTTATTTTGGTACC	1140
Dp	1086	TACTTAAACTCAGTTAATCTTTATATCATTCATTTGGTGATAGTATTTATTTTGGTACT	1145

QY	1141	TCGTGAATTCACATCATTTATGGTACTATTTCATTTCCCACTAGATTATCTGATCTAACCGCT	1200
Db	1146	TCCTGAATTTACCATCATTTATGTATGACTATTTCATTTCCCACTAGATTATCTGATCTAACCGCT	1205
QY	1201	TTGGAATCAAGTTAAACAGATGATATTTGAAGCTTCAAGCACTGATPAACGGGTACAACA	1260
QY	1261	AACGGAACAAGACAACAGCTAGATACATCTAGTGGTTCACAGGCTGGCGAACAAGAAAT	1320
Db	1266	AACGGAACAAGACAACAACAGTATACATCTAGTGGTTCACAGGCTGGCGAACAAGAAAT	1325
QY	1321	ACTACTAACACTTCTCAACAGTTCCTATATCTTAAATTAATCTATTCGTAGTTTTGGA	1380
Db	1326	ACTACTAACACTTCTCAACAGTTCCTATATCTTAAATTAATCTATTCGTAGTTTTGGA	1385
QY	1381	ATTGATGTAATCAACACTCTCTGCAACAACAAATAGTGAACATAATTTGGGCGAGATCTTAC	1440
Db	1386	ATTGATGTAATCAACACTCTCTGCAACAACAAATAGTGAACATAATTTGGGCGAGATCTTAC	1445
QY	1441	GTTATTGGAAGCAAGAAATATATGCTGAATACAGATTAGTATTGATTTCAAAATGAATTTCAATA	1500
Db	1446	GTTATTGGAAGCAAGAAATATATGCTGAATACAGATTAGTATTGATTTCAAAATGAATTTCAATA	1505
QY	1501	ACTATATCGAGAAACCTTTATCCGAACAACAATTTGGTGTGTTGGTTTTACTTCAACAGCT	1560
Db	1506	ACTATATCGAGAAACCTTTATCCGAACAACAATTTGGTGTGTTGGTTTTACTTCAACAGCT	1565
QY	1561	TTCAAGAGATGTTTAAAGAGCTTCCTTATPAAGGGTATCCAAAGCTCCAAAGTGGAAACCTCCAA	1620
Db	1566	TTCAAGAGATGTTTAAAGAGCTTCCTTATPAAGGGTATCCAAAGCTCCAAAGTGGAAACCTCCAA	1625
QY	1621	CCTTTCTCTATACGATTTTGGTTATTATAGATACCAACAACAATCTAGAACAGAACTTTCGG	1680
Db	1626	CCTTTCTCTATACGATTTTGGTTATTATAGATACCAACAACAATCTAGAACAGAACTTTCGGA	1685
QY	1681	TACGGAAACATATAGCTTTTAAACAACAAGCCCTTACGAGCTATAGATCTTCCAGAGTA	1740
Db	1686	TACGGAAACATATAGCTTTTAAACAACAAGCCCTTACGAGCTATAGATCTTCCAGAGTA	1745
QY	1741	GSTACTGAACCAATCAATTTTGAAGAAGCTTCATTAACATCCGCTTATGSGGAGAT	1800
Db	1746	GSTACTGAACCAATCAATTTTGAAGAAGCTTCATTAACATCCGCTTATGSGGAGAT	1805
QY	1801	CTAACTGAGAAGGCTGAGAAAGTTTCTATATCTCCATCTTATPAAGAGCAAGGTGAC	1860
Db	1806	CTAACTGAGAAGGCTGAGAAAGTTTCTATATCTCCATCTTATPAAGAGCAAGGTGAC	1865
QY	1861	ACACCAGAAACCCGAAGCATCTTCCAAATCTGGCTATTCTGATTAATCTATTAGACTAT	1920
Db	1866	ACACCAGAAACCCGAAGCATCTTCCAAATCTGGCTATTCTGATTAATCTATTAGAGTAT	1925
QY	1921	CAATCAGTTTATAGATTTGATGAGATTTGAANAATTACTTAAATGTTGGGGTTAAAGCATCA	1980
Db	1926	CAATCAGTTTATAGATTTGATGAGATTTGAANAATTACTTAAATGTTGGGGTTAAAGCATCA	1985
QY	1981	AGCTTCTTAAACTCAAAATAGACAACAATCCAAACGGCTAGAGAATGATTGGCTGCAACATCA	2040
Db	1986	AGCTTCTTAAACTCAAAATAGACAACAATCCAAACGGCTAGAGAATGATTGGCTGCAACATCA	2045
QY	2041	TACTTAAAGTCACAAAATTTGGATTAGCTAGAACATCTGATTTACCAACCAACAACCATTC	2100
Db	2046	TACTTAAAGTCACAAAATTTGGATTAGCTAGAACATCTGATTTACCAACCAACAACCATTC	2105
QY	2101	GGAAACAATCAACCAAGTTATTTTCAATCACCCTGGTGATCAGTTCTCATCAATTPAGAT	2160
Db	2106	GGAAACAATCAACCAAGTTATTTTCAATCACCCTGGTGATCAGTTCTCATCAATTPAGAT	2165
QY	2161	ATTGAACAACAATCTCCCTGGTAAACAGTTATGATGACTCTTATATCCAAAGAANAATAT	2220
Db	2166	ATTGAACAACAATCTCCCTGGTAAACAGTTATGATGACTCTTATATCCAAAGAANAATAT	2225
QY	2221	AAATCTAGTGTATTATACATTTAAAGATTACTGATCTCACTAATTAACCCGATGAGCTCAAGCTCA	2280

Db	2226	AAATCTAGAGTTTATACATTAAAGNTAGCTSACTCAAGTAACCTGATCGTCAAGTCA	2285
Qy	2281	TTTCAGTCCAAACAGTTTAAATGACGTTAATGAATTCGTGTAATCTTACCTTATATAGC	2340
Db	2286	TTTCAGTCCAAACAGTTTAAATGACGTTAATGAATTCGTGTAATCTTACCTTATATAGC	2345
Qy	2341	AATTCATCTCTATACAGTAATAGCTGCTAGTAAATGTTTCATTTGTTTCATCAACCCCTGGT	2400
Db	2346	AATTCATCTCTATACAGTAATAGCTGCTAGTAAATGTTTCATTTGTTTCATCAACCCCTGGT	2405
Qy	2401	TTCTCCGATCATATATCTGCTGTAAATATCATTTAATCAAGAACTATATGATATGCTTTT	2460
Db	2406	TTCTCCGATCATATATCTGCTGTAAATATCATTTAATCAAGAACTATATGATATGCTTTT	2465
Qy	2461	GAAGGTCCTGCTGCTAGTATATCATCTGATTTCTGGGACAAATCCAAATCCAAACCCGAT	2520
Db	2466	GAAGGTCCTGCTGCTAGTATATCATCTGATTTCTGGGACAAATCCAAATCCAAACCCGAT	2525
Qy	2521	GAGTACTTAATTCAAAATGGGCTTCATCATCAAGGGCTAGAGAACTTGCTTACAAACCA	2580
Db	2526	GAGTACTTAATTCAAAATGGGCTTCATCATCAAGGGCTAGAGAACTTGCTTACAAACCA	2585
Qy	2581	AGCTTCTTAAACAGTTAGTTAGTCACTTCCTGCTAATGCTGTGTAATCTTACCTACCGTGA	2640
Db	2586	AGCTTCTTAAACAGTTAGTTAGTCACTTCCTGCTAATGCTGTGTAATCTTACCTACCGTGA	2645
Qy	2641	GTCGTGATCTCTGATGTAATTTAAACAAACCAAAACCTACCTCTTAAAGTTCCAGTCCAA	2700
Db	2646	GTCGTGATCTCTGATGTAATTTAAACAAACCAAAACCTACCTCTTAAAGTTCCAGTCCAA	2705
Qy	2701	TACTTAGATGTAATATATGATGCTAAATTTAAAGAACAAATATTTAGTACATTTCTCT	2760
Db	2706	TACTTAGATGTAATATATGATGCTAAATTTAAAGAACAAATATTTAGTACATTTCTCT	2765
Qy	2761	TATTAACAATTTGGGCGCTTACCTTCATGGTAGTGCTTACAGCAATTTGTAATGATTA	2820
Db	2766	TATTAACAATTTGGGCGCTTACCTTCATGGTAGTGCTTACAGCAATTTGTAATGATTA	2825
Qy	2821	GGTATTTCTTGCAATTAATGATCATCTTACAGATTAGCTATCGGTATTCCTTTAGAGCTCAA	2880
Db	2826	GGTATTTCTTGCAATTAATGATCATCTTACAGATTAGCTATCGGTATTCCTTTAGAGCTCAA	2885
Qy	2881	AGAAATTTCAAGACAAAGGGTTCAAAACAATTCAAAAGTTGATACCTGTAAGTCT	2940
Db	2886	AGAAATTTCAAGACAAAGGGTTCAAAACAATTCAAAAGTTGATACCTGTAAGTCT	2945
Qy	2941	GCTGTGTTGTTCACTTTACAAAGAAAGTATTTACCCAAATCGTATACGTTAAGAAAAACCT	3000
Db	2946	GCTGTGTTGTTCACTTTACAAAGAAAGTATTTACCCAAATCGTATACGTTAAGAAAAACCT	3005
Qy	3001	GCTGCTTTAGGTGCTGTGTAATCTGGTGATTAAGAAACCTGCTGCTGCTGCTTAAACCGCT	3060
Db	3006	GCTGCTTTAGGTGCTGTGTAATCTGGTGATTAAGAAACCTGCTGCTGCTGCTTAAACCGCT	3065
Qy	3061	GCTCAGCTTAAACCATCTGCAACCAAGCTAGCTCACAAGCTTAAACCAACTGGGCTTAA	3120
Db	3066	GCTCAGCTTAAACCATCTGCAACCAAGCTAGCTCACAAGCTTAAACCAACTGGGCTTAA	3125
Qy	3121	TCTGTGTCGCTTACAAAACCACTGCTCTAGCCAGCTGCTCCAAAACCAACGCGCTCC	3180
Db	3126	TCTGTGTCGCTTACAAAACCACTGCTCTAGCCAGCTGCTCCAAAACCAACGCGCTCC	3185
Qy	3181	AAAGAAATTA 3189	
Db	3186	AAAGAAATTA 3194	
RESULT 6			
AX665242		3129 bp	DNA
LOCUS			linear
DEFINITION	Sequence 79 from Patent EPI275716.		PAT 26-MAR-2003
ACCESSION	AX665242		

[illegible]

|||||
Db 725 CTCAACTAAACCTACTGATTAATCAAAATCTTCACTTTAATTCAGAGCAATCG 784
845 CTGGCTGCAAAACATAGATCGATTCATATTTGAATGCAAGCTACAAATTAACATCTT 904
785 CTGGTGCANACATATATCGATTCATATTTGAATGCAAGCTACAAATTAACATCTT 844
905 TCCAAATAGATGAAAAATTTGTTATTCAGAAATGAGCTGTTCTGAAGAAATAAAAAT 964
845 TCCAAATAGATGAAAAATTTGTTATTCAGAAATGAGCTGTTCTGAAGAAATAAAAAT 904
965 TTACAAGATTAGCTACTGAGAGTTGCCAAGCAACAAAGATTTGATTCCTTGACATAC 964
905 TTACAAGATTAGCTACTGAGAGTTGCCAAGCAACAAAGATTTGATTCCTTGACATAC 964
1025 CCGGAGCTCCACAGATTACTTTAAAGAGATTCAGTTAAGCTTTTCAAGCTTACT 1084
965 CCGGAGCTCCACAGATTACTTTAAAGAGATTCAGTTAAGCTTTTCAAGCTTACT 1024
1085 TAACTCAGTTAATTCCTTATATCATTCATTCATTCATTCATTCATTCATTCATTC 1144
1025 TAACTCAGTTAATTCCTTATATCATTCATTCATTCATTCATTCATTCATTCATTC 1084
1145 AATTACCATCATTTATGTTACTATTCATTCCTCAACTGATTTATCTGATTCACCGCTTGA 1204
1085 AATTACCATCATTTATGTTACTATTCATTCCTCAACTGATTTATCTGATTCACCGCTTGA 1144
1205 ATCAAGTTAAACAGATGATTAATGAAGCTTCAAGCACTGATTAAGCTTACAAACAAC 1264
1145 ATCAAGTTAAACAGATGATTAATGAAGCTTCAAGCACTGATTAAGCTTACAAACAAC 1204
1265 GAACAACGACACAGCTGATACATCTAGTGTTCACACAGTGTGGAACAGAAATAC 1324
1205 GAACAACGACACACAGCTGATACATCTAGTGTTCACACAGTGTGGAACAGAAATAC 1264
1325 CTACACTCTCTCAACAGTTTCTAATCTTAACTTAAATCTTAACTTAACTTAACTTGA 1384
1265 CTACACTCTCTCAACAGTTTCTAATCTTAACTTAACTTAACTTAACTTAACTTGA 1324
1385 ATATGATTAACCACTTCTGCAACCAAAATGATGAACTAATTTGGAGATCTTAACTG 1444
1325 ATATGATTAACCACTTCTGCAACCAAAATGATGAACTAATTTGGAGATCTTAACTG 1384
1445 TTGAACCAAGATATATGCTGATACAGATTAGTATTCAAAATGAAATCCAAATAC 1504
1385 TTGAACCAAGATATATGCTGATACAGATTAGTATTCAAAATGAAATCCAAATAC 1444
1505 ATGACAGAAACCTTTATCCGAAACAAATGAGTGTGTTGTTTACTTCAACAGTTCAA 1564
1445 ATGACAGAAACCTTTATCCGAAACAAATGAGTGTGTTGTTTACTTCAACAGTTCAA 1504
1565 GAGTACTTTTAAAGAGCTTCTTAAAGGAGATCAAGCTGCAACTGGAAATCTCAACCTT 1624
1505 GAGTACTTTTAAAGAGCTTCTTAAAGGAGATCAAGCTGCAACTGGAAATCTCAACCTT 1564
1625 TCTTATACGATATTGCTTATTTAGATACCAACAACTAGAACAGAACTTTCTGTAGC 1684
1565 TCTTATACGATATTGCTTATTTAGATACCAACAACTAGAACAGAACTTTCTGTAGC 1624
1685 GAACATATAGCTTTTAAACACAGCCCTTACAGACGATTTAGATTCCTCAAGATAG 1744
1625 GAACATATAGCTTTTAAACACAGCCCTTACAGACGATTTAGATTCCTCAAGATAG 1684
1745 CTGAAGCAATCAATTTAGAGAACTTCAATACATACCTGTTTAAAGGAGATTA 1804
1685 CTGAAGCAATCAATTTAGAGAACTTCAATACATACCTGTTTAAAGGAGATTA 1744
1805 CTGAAGCAATCAATTTAGAGAACTTCAATACATACCTGTTTAAAGGAGATTA 1804
1745 CTGAAGCAATCAATTTAGAGAACTTCAATACATACCTGTTTAAAGGAGATTA 1804
1865 CAGAAAGCGAAGCATCTTCAATCTGCTAATCTGATTAATCTGATTAATGATTAAT 1924
|||||

Db 1805 CAGAAAGCGAAGCATCTTCAATCTGCTAATCTGATTAATCTGATTAATGATTAAT 1864
1925 CAGTTTAAAGATTTGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 1984
1865 CAGTTTAAAGATTTGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 1924
1985 TCTTAAACCTCAATATAGACCAAAATCCAAAGGCTTACAAATGATTTGCTGCAACAT 2044
1925 TCTTAAACCTCAATATAGACCAAAATCCAAAGGCTTACAAATGATTTGCTGCAACAT 1984
2045 TTAAGATCACAATTTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2104
1985 TTAAGATCACAATTTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2044
2105 CAACCTCACAAGTTATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2164
2045 CAACCTCACAAGTTATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2104
2165 GAACAATCTTCCCTGCTAACCACTTATGATTAATGATTAATGATTAATGATTAATGATTA 2224
2105 GAACAATCTTCCCTGCTAACCACTTATGATTAATGATTAATGATTAATGATTAATGATTA 2164
2225 CTAGTGTATATACATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2284
2165 CTAGTGTATATACATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2224
2285 GTCCAAACAAGTTAATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2344
2225 GTCCAAACAAGTTAATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2284
2345 CATTCTATACAGTAATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2404
2285 CATTCTATACAGTAATTTAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2344
2405 CTGATCATATATCTGCTGTAATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2464
2345 CTGATCATATATCTGCTGTAATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2404
2465 GTTCTGTGTATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2524
2405 GTTCTGTGTATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2464
2525 ACTTAATTCAAAATGAGGCTTACATGATTAATGATTAATGATTAATGATTAATGATTA 2584
2465 ACTTAATTCAAAATGAGGCTTACATGATTAATGATTAATGATTAATGATTAATGATTA 2524
2585 TCTTAAACAGTTAGTTAGTACTTACATGATTAATGATTAATGATTAATGATTAATGATTA 2644
2525 TCTTAAACAGTTAGTTAGTACTTACATGATTAATGATTAATGATTAATGATTAATGATTA 2584
2645 TTGATCCTGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2704
2585 TTGATCCTGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2644
2705 TAGATGTAAGTATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2764
2645 TAGATGTAAGTATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2704
2765 ACAACCTTTGGCGCTTACCTTACCTGATTAATGATTAATGATTAATGATTAATGATTA 2824
2705 ACAACCTTTGGCGCTTACCTTACCTGATTAATGATTAATGATTAATGATTAATGATTA 2764
2825 TTCTGCAATTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2884
2765 TTCTGCAATTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2824
2885 AATTACAAGCAAAAGGCTTCAAAACCAATTTCAAAAGGTTGATTAATGATTAATGATTAAT 2944
2825 AATTACAAGCAAAAGGCTTCAAAACCAATTTCAAAAGGTTGATTAATGATTAATGATTAAT 2884
2945 TTGCTGATGATTTCAAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 3004
2885 TTGCTGATGATTTCAAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2944

QY 3005 CTTTAGGTCGTGTAATCTGTGTGATTAAGAAACCTGCTGCTGCTAAACCTGCTC 3064
 |||||||
 Db 2945 CTTTAGGTCGTGTAATCTGTGTGATTAAGAAACCTGCTGCTGCTAAACCTGCTC 3004
 |||||||
 QY 3065 CAGCTAAACATCTGCACCAAAAGCTAGCTACACGCTAAACCAACGCGGCTAAATCTG 3124
 |||||||
 Db 3005 CAGCTAAACATCTGCACCAAAAGCTAGCTACACGCTAAACCAACGCGGCTAAATCTG 3064
 |||||||
 QY 3125 GTGCGGCTACAAACCACTGCTCTTAAGCCAGCTGCTCAAAACCAACGCTCCCAAG 3184
 |||||||
 Db 3065 GTGCGGCTACAAACCACTGCTCTTAAGCCAGCTGCTCAAAACCAACGCTCCCAAG 3124
 |||||||
 QY 3185 AATAA 3189
 |||||
 Db 3125 AATAA 3129
 |||||
 RESULT 7
 AF083976 7141 bp DNA linear BCT 22-AUG-1998
 LOCUS Mycoplasma gallisepticum adherence protein gene, complete cds.
 DEFINITION AF083976
 ACCESSION AF083976
 VERSION AF083976.1 GI:3450890
 KEYWORDS
 ORGANISM Mycoplasma gallisepticum
 Mycoplasma gallisepticum
 Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 REFERENCE 1 (bases 1 to 7141)
 AUTHORS Goh, M.S., Gorton, I.S., Forsyth, M.H., Troy, K.E. and Geary, S.J.
 TITLE Molecular and Biochemical Analysis of a 105 kDa Mycoplasma
 gallisepticum cytoadhesin
 JOURNAL Microbiology (1998) In press
 REFERENCE 2 (bases 1 to 7141)
 AUTHORS Goh, M.S. and Geary, S.J.
 TITLE Mycoplasma gallisepticum adherence protein gene (gapA) and flanking
 regions
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 7141)
 AUTHORS Goh, M.S. and Geary, S.J.
 TITLE Direct Submission
 JOURNAL Submitted (13-AUG-1998) Pathobiology, University of Connecticut, 61
 North Eagleville Road, Storrs, CT 06269, USA
 FEATURES
 source
 1..7141
 /organism="Mycoplasma gallisepticum"
 /mol_type="genomic DNA"
 /strain="S6"
 /db_xref="taxon:2096"
 2994..5888
 /note="GapA: Mycoplasma pneumoniae P1 homolog; gapA;
 formerly referred to as Mgp1"
 /codon_start=1
 /transl_table=4
 /product="adherence protein"
 /protein_id="AAC32609.1"
 /db_xref="GI:3450891"
 /translation="MQRFNKRLKIKLTKTLCLMHLVLAIKVLQILLVVOY
 VILEVLLIPASSTARLVEEGNSAGMTPASTSEVIDYNDOKIKRITLD
 SSSSSSINGRTYANINTOKNGIYVENVNENSEPEVEMAKIKRWIND
 PSTWGSANGKATVLOFYKIDNPVANNRBAKYRKLETOTTPLLIDSSPY
 EHEPEWEGNOFVPMWYITNIGGLIAKGMWYILFGNGTWNNSALSIVFRKE
 NRTAEAGNTRIVGYPYGIILSLISFDATRNGIALAAPLIGDGYHVPRLAVGS
 SPRGANGNIFGSAITWGTNGNFIDTKHSPVIEDATFTITVNSSGVLANSSQO
 STSPMPNSNGNESIPYRMVNSYDYSVFAALISKPAQNTKOVESLPTAKIDL
 NSLPKTKTOENNIFFSAMLDGRMSLIGTKRDSMTLTNTINNTYNTQOOLASTAG
 ENANPRILIALTTAGFDRDROIGNVVYYSNNTKFTYVYVGVGATITPEVOYRK
 TSANITVILFRIDFGSTPATDANTVSGSKINGAVLSTGOOGYNGSIYKKA
 TPSQGYTQWDFGLTTASNAVISMTAAGSYLRDDDTVSYSKIPKRTAAVN
 VRSLSYVQINGESTVNTAVRSPSSALILMPKRTIPLMANRDNVVGOGAFISRD
 IPSSEFNKINDIYTTAEADTEVLDSKYSINSIYRTIPPNPHIRLRLVIDRSATN
 DFILPOVLVDGEYVAVPOANSVFVSDQFTGFDALPGVILPAVLSITITIALALA
 LGIGITGPMQNRMLKQFAISNKKVDILTAVGVSKFOIINRSVTIKTPQMLQ

BASE COUNT 2506 a 1261 c 1263 g 2100 t 11 others
 ANKDGASSPEKSPAPAKKPTGPTKSPAGTAPAKKAPAPKIE"
 ORIGIN
 Query Match 33.0%; Score 1052.4; DB 1; Length 7141;
 Best Local Similarity 92.6%; Pred. No. 1.2e-167;
 Matches 1145; Conservative 4; Mismatches 81; Indels 6; Gaps 4;
 QY 1 ATGAATATTTCTAAAAAATTAAAGTATATCATTTAGTAGGTGATAGCTGATTTTGA 60
 Db 5910 ATGAATATTTCTAAAAAATTAAAGTATATCATTTAGTAGGTGATAGCTGATTTTGA 5969
 |||||||
 QY 61 GCTCTGTTCTGCAAGCTTTGCTTTAAGCAATCAGATAAGTAAAGTAAACGCCAA 120
 Db 5970 ACTCTGTTCTGCAAGCTTTGCTTTAAGCAATCAGATAAGTAAAGTAAACGCCAA 6029
 |||||||
 QY 121 TTAGTTAATCAAGCAAGCAAGCTAGATGCTATTTCTTTAGACTGTCAGTCTTGACAA 180
 Db 6030 TTAGTTAATCAAGCAAGCAAGCTAGATGCTATTTCTTTAGACTGTCAGTCTTGACAA 6089
 |||||||
 QY 181 AATGATTCGTTGTTCAATACAGTTCTTTAGAGATGTTGATGATACCTTTATACAGCACT 240
 Db 6090 AATGATTCGTTGTTCAATACAGTTCTTTAGAGATGTTGATGATACCTTTATACAGCACT 6149
 |||||||
 QY 241 AATGAAACAATTTATCAATTAAGTATTTTACTTAAACCTATATAGTTTATAGTCTAAGT 300
 Db 6150 AATGAAACAATTTATCAATTAAGTATTTTACTTAAACCTATATAGTTTATAGTCTAAGT 6209
 |||||||
 QY 301 GATGATTTGCTGATGATCAAAAGTAAACAATATGTTTCAAGTATTCACAACTGCGCAAT 360
 Db 6210 GATGATTTGCTGATGATCAAAAGTAAACAATATGTTTCAAGTATTCACAACTGCGCAAT 6269
 |||||||
 QY 361 AGATTTGATCAAGCAAGCAAGCAAGCAAGTATATGCTCTGTTGTTAATGATGAGCTAAC 420
 Db 6270 AGATTTGATCAAGCAAGCAAGCAAGCAAGTATATGCTCTGTTGTTAATGATGAGCTAAC 6329
 |||||||
 QY 421 GTTCAATTAATAAAGAAATTAATTAATCAATTAAGTATGTTGTTAATAAGCAATTC 480
 Db 6330 GTTCAATTAATAAAGAAATTAATTAATCAATTAAGTATGTTGTTAATAAGCAATTC 6389
 |||||||
 QY 481 AAGTTTGAATTTGCTGCTGTTGATTAATCAAGCTAGCTAATTTGATTTACGATGAG 540
 Db 6390 AAGTTTGAATTTGCTGCTGTTGATTAATCAAGCTAGCTAATTTGATTTACGATGAG 6449
 |||||||
 QY 541 ACTAATATTAATTTTACAAACCAAGCTCAAGTGAATTTGA--TGACTCTATTTAGA 599
 Db 6450 ACTAATATTAATTTTACAAACCAAGCTCAAGTGAATTTGA--TGACTCTATTTAGA 6509
 |||||||
 QY 600 TGCGCCAACTTCTACTTAAAGATTTACACGAGTTGGATTAATTAATCAATTC 659
 Db 6510 TGCGCCAACTTCTACTTAAAGATTTACACGAGTTGGATTAATTAATCAATTC 6599
 |||||||
 QY 660 GATCTTACCAATGAACGTCACACGAGTTGTTCCCTTGCCAGTAGTAGTACTAGTGG 719
 Db 6570 GATCTTACCAATGAACGTCACACGAGTTGTTCCCTTGCCAGTAGTAGTACTAGTGG 6629
 |||||||
 QY 720 AACAAATGCT-GATGATGGGATGTTGATGTTGGGAATGGCAATTAATTAATCAAGT 778
 Db 6630 TAATACGCTTACAGGTGGACATTTGATTTGCGAAGGGAATGGCTCTCTAATGATC 6689
 |||||||
 QY 779 CTTATGCTAAACATAAACCACACTAGTATCAATCAATCTTCAACTTTAATTCAGAG 838
 Db 6690 CATTATGACGCTTAAAGCGTTTCAAAAT---AMCCCTCAAGTTCAATTCAGAG 6746
 |||||||
 QY 839 CAATGCTGCTGCAACCAATAGATACGATTTCTCAATGGAATTCAGATAGATTAATAA 898
 Db 6747 CTAACCTGCGCAACCAATAGATAGATTTCTCAATGGAATTCAGATAGATTAATAA 6806
 |||||||
 QY 899 CATTTTCAATTAAGATTAATAATTTGTTATCCAGATGCACTGTTTGAAGGAATA 958
 Db 6807 CATGTTCAATTAAGATTAATAATTTGTTATCCAGATGCACTGTTTGAAGGAATA 6866
 |||||||
 QY 959 AAATATTTACAAAGATTAAGTACTGGAAGTTTGCCAAAGCAACGAAGATTTGGATTTCTTG 1018

Db 6867 AAAATATACAGATAGCTAGCTGGAAGTTGGCAACAGAGATATGAATCTTG 6926
QY 1019 ACATACCCGGGACTCCACAAGTTACTTTAAAGAAATTCAGTTAATTTTCAAGC 1078
Db 6927 ACATACAGGAGCTCCAGAAGTTACTTTAAAGAAATTCAGTTAATTTTCAAGC 6986
QY 1079 TATACCTAACACAGTATGTTTCTTATCATTCCTGTTGATAGTATTTATTTTGGTA 1138
Db 6987 TATACCTAACACAGTATGTTTCTTATCATTCCTGTTGATAGTATTTATTTTGGTA 7046
QY 1139 CCTGTAATTAACCATTAATGTTACTATTCATTCCTCAACATGATTTATGATCAACG 1198
Db 7047 CTCTGTAATTAACCATTAATGTTACTATTCATTCCTCAACATGATTTATGATCAACG 7106
QY 1199 CTCTGTAATTAACCATTAATGTTACTATTCATTCCTCAACATGATTTATGATCAACG 1234
Db 7107 CTCTGTAATTAACCATTAATGTTACTATTCATTCCTCAACATGATTTATGATCAACG 7141

RESULT 8
AB033210 1131 bp DNA linear BCT 07-APR-2000
LOCUS Mycoplasma gallisepticum gene for 120-kDa membrane protein MG3,
DEFINITION partial cds, strain:56.
ACCESSION AB033210
VERSION AB033210.1 GI:7527370
KEYWORDS 120-kDa membrane protein MG3.
SOURCE Mycoplasma gallisepticum
ORGANISM Mycoplasma gallisepticum
REFERENCE 1 (bases 1 to 1131)
AUTHORS Yoshida,S.
TITLE Mycoplasma gallisepticum S6-strain gene encoding a 120-kDa membrane protein
JOURNAL Published Only in Database (2000)
AUTHORS Yoshida,S.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Shigeto Yoshida, Jichi Medical School,
Department of Medical Zoology, Yakushiji 3311-1,
Minamikawachimachi, Tochigi 329-0498, Japan
(E-mail:shigeto@jichi.ac.jp, Tel:81-285-58-7339,
Fax:81-285-44-6489)

FEATURES
source
1. 1131
Location/Qualifiers
/organism="Mycoplasma gallisepticum"
/mol_type="genomic DNA"
/strain="S6"
/db_xref="taxon:2096"
1. 1131
/gene="mgc3"
1. >1131
/gene="mgc3"
/codon_start=1
/transl_table=4
/product="120-kDa membrane protein MG3"
/protein_id="BA94277.1"
/db_xref="GI:7527371"
/translation="MNISKKLKSYLLIGLAVFGLGSAFGRKQSDKSDNTQLVNO
ARTLDANSVRLAGLGGSLFNTLRVDVDFNTAANGTILIKLDSSTKPYGLDLDSD
FAGYKQKIVSDYTSRNRFDQROTARAYALLVNDANVHLKRIINNSNIGRNNS
KFVIGVDNPAHYIRFTDGTKEFTKOTGEIVDFILDAPILPKDLHDWYNLITO
RKLIPDVNTAVPMPVGRVGSNSATDGTGPDGNGMPSANDPIATSKSSQNNPSSFN
SSAOPCTNNRKYSEVNVKRIKTSFOLDKFTYPTWNTGSEENKNTIRLATGSLPNEG
YWLIDIPGPEVTLKEDSVNFSRLYLNSVNSLSFTGDSLYIF"

BASE COUNT 411 a 169 c 197 g 354 t
ORIGIN
Query Match 31.1%; Score 993.2; DB 1; Length 1131;
Best Local Similarity 92.9%; Pred. No. 1.7e-157;
Matches 1053; Conservative 0; Mismatches 78; Indels 3; Gaps 1;

QY 1 ATGAATATTCTTAAAAAAGCTTAAAGTTATACATGATAGTGATGATCTGATTTGCA 60
Db 1 ATGAATATTCTTAAAAAAGCTTAAAGTTATACATGATAGTGATGATCTGATTTGCA 60
QY 61 GCTCTGTTGTTGCAAGCTTTGGCTTTAAGCAATCAGATTAAGTAACGATTAACGCA 120
Db 61 ACTCTGTTGTTGCAAGCTTTGGCTTTAAGCAATCAGATTAAGTAACGATTAACGCA 120
QY 121 TTATGTAATCAACCAAGACGCTAGATGCTAATTCGTGTATAGACTGCAGCTTGACAA 180
Db 121 TTATGTAATCAACCAAGACGCTAGATGCTAATTCGTGTATAGACTGCAGCTTGACAA 180
QY 181 AATGTTGCTGTTTCAATACAGTCTTAGAGATGTTGATGATTAACCTTAACGAGCT 240
Db 181 AATGTTGCTGTTTCAATACAGTCTTAGAGATGTTGATGATTAACCTTAACGAGCT 240
QY 241 AATGGAACAATTAATCAATAGATGTTTACTAAACCATTAATAGTGTAGATCTAGT 300
Db 241 AATGGAACAATTAATCAATAGATGTTTACTAAACCATTAATAGTGTAGATCTAGT 300
QY 301 GATGATGTTGCTGATACCAAGTAAACCAATAGTTTCAGATTAACCAACTAGCAGAA 360
Db 301 GATGATGTTGCTGATACCAAGTAAACCAATAGTTTCAGATTAACCAACTAGCAGAA 360
QY 361 AGATTGATCAACCAAGACCAAGCATATTAATGCTGTTGTTAATGATGAAGCTAAC 420
Db 361 AGATTGATCAACCAAGACCAAGCATATTAATGCTGTTGTTAATGATGAAGCTAAC 420
QY 421 GTTCATTTAAAAAGAAATTAATACATCAATAGAATGATGTAATGAACCAATTC 480
Db 421 GTTCATTTAAAAAGAAATTAATACATCAATAGAATGATGTAATGAACCAATTC 480
QY 481 AAGTTGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Db 481 AAGTTGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
QY 541 ACTAAATTTAATTTTCAACCAACCAAGCAAGGTAATGTTATGATCACTTATTAAGAT 600
Db 541 ACTAAATTTAATTTTCAACCAACCAAGCAAGGTAATGTTATGATCACTTATTAAGAT 600
QY 601 GCGCCAACTTTACCTTAAGATTTTACACCCAGATGTTATTAATCAATCAAGCAAG 660
Db 601 GCGCCAACTTTACCTTAAGATTTTACACCCAGATGTTATTAATCAATCAAGCAAG 660
QY 661 ATCTTACCAATGACGCAACATGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720
Db 661 ATCTTACCAATGACGCAACATGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720
QY 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 ATAGCAACGCTTAAAGCGGCTTCAAAAT---AACCCCAAGTTCAATTCAGCGCT 837
Db 781 ATAGCAACGCTTAAAGCGGCTTCAAAAT---AACCCCAAGTTCAATTCAGCGCT 837
QY 841 ATGCGCTGTCGAACCAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 ATGCGCTGTCGAACCAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 898 TCGTTTCAATTAACGCAAAATTTGTTTATTCGAATGACCTGTTCTGAAAGCAATTA 957
Db 898 TCGTTTCAATTAACGCAAAATTTGTTTATTCGAATGACCTGTTCTGAAAGCAATTA 957
QY 961 AATATTACAAGATTTAGCTAGTGAAGTTTCCCAAGCAAGCAAGATATTGATCTTGAC 1020
Db 961 AATATTACAAGATTTAGCTAGTGAAGTTTCCCAAGCAAGCAAGATATTGATCTTGAC 1020
QY 958 AATATTACAAGATTTAGCTAGTGAAGTTTCCCAAGCAAGCAAGATATTGATCTTGAC 1017
Db 958 AATATTACAAGATTTAGCTAGTGAAGTTTCCCAAGCAAGCAAGATATTGATCTTGAC 1017
QY 1021 ATACCGGGAGCTGCAAGATTTACTTTAAAGAAATTCAGTTAAGCTATTTCAGACTA 1080
Db 1021 ATACCGGGAGCTGCAAGATTTACTTTAAAGAAATTCAGTTAAGCTATTTCAGACTA 1080
QY 1018 ATACCGAGGAGCTGCAAGATTTACTTTAAAGAAATTCAGTTAAGCTATTTCAGACTA 1077
Db 1018 ATACCGAGGAGCTGCAAGATTTACTTTAAAGAAATTCAGTTAAGCTATTTCAGACTA 1077
QY 1081 TACTTAAATCACTTAATTTCTTATATCAATTCATTTGATGATTAATTAATTTT 1134
Db 1081 TACTTAAATCACTTAATTTCTTATATCAATTCATTTGATGATTAATTAATTTT 1134

TITLE Nucleotide sequence of the MgpA (mgp) operon of Mycoplasma genitalium and comparison to the Pl (mp) operon of Mycoplasma pneumoniae

JOURNAL Gene 82 (2), 259-267 (1989)

MEDLINE 9060815

PUBMED 2583522

COMMENT Original source text: M.genitalium (strain G-37, ATCC 33530) DNA.

FEATURES

source location/Qualifiers

1. 8760

/organism="Mycoplasma genitalium"

/mol_type="genomic DNA"

/db_xref="taxon:2097"

298..1059

/note="29 kd protein"

/codon_start=1

/transl_table=4

/protein_id="AA25419.1"

/db_xref="GI:150158"

/translation="MTFTQIDSEFLKEALIVDANKRIELRELDKNEFLAVIRIDHNNEDLNTSFNVEESVYACCEQIVEMATVAKWTPVPAATLLYIGITDSNRLKXNTSYRTLYLAALLYKAKADIRIYHDLNNTSLADLKFKYVYVNHFTQGVYFEICTKTIQKRLMTADOCARVNLSTADIKIMLEFTEQANNEIRIELRSGINVRDAIKYGGGGHNNASALITTKQISDVVDCVKKIYVN"

1066..5400

/note="attachment protein MgpA"

/codon_start=1

/transl_table=4

/protein_id="AA25420.1"

/db_xref="GI:150159"

/translation="MHQPKRLAKKSMALTAALTGLVITGVGYELFNQNRKSSVSIFYQPKLQSVKHOQAVDELTPTMTNNNNFSLKTGPNPSFGLVRSQNDNLNLSYKSSDDMLKYNVEXYLDGQONFALIRYDNNNGALDILAKENNSTYORINGEDTPPFRGFLTGNAPTDMMFKIGVPEVYVOSPSPLVLLVRYKALRYHNINQYVKESELYKATNSPNTQRIQKDSKPKVDSKQKESLSETIASSSSMASTRAKALKEVERGSDSLKNDPAKPLAKHNSGKLEAEKEFEKAMPPLTLQIQAREKMGATVVSFYDAPYSENHATFGLVDHIDRKAVENPPSKMPKMHIMYNAENLLOQTGVFNPBRHPEWFDGAKADNTSPGFVQVDHKKDGFKNSSSPIALPEAYFANIGMVAIGNSVPIFGNGHATKMTNPISIGVRIKYTDNFSKSSVTGMPYAVLFGGLNPOTNGKLDPLGTNRMVEYPRVAVGKVGQVLVAGLTMDGNATVRLKYDQLEHLNVAQOGILREDLOITFTPGMARNPDIPIVGAALODMGSKRPHEYLINPDIDONVNDTVEALISSYKNTDKIKHYIPRISGLIAMQLEFMSNKLITNPISANFVNENSTAPSLFAALINEDLTLGLSDKIFYGKEEAEAEADRFNQLSLNPNPTNMAARYLVNVOFRFTTGPNLDSSTFOQDFLFIWIGNKRPNSPSPSSASSSTPLPTNINAVGKSMITQHLNKNENTRMVFIIPNESPDIWAGAGYVSAQKNGIPEQVKSNSITPEPDSNDKVTSPGSGSKPTTYPALPSPISPTSPMIALFTNNNNORNDILRSLGTTIPLINKSGSDNDKNDSEOKMDETPNEGLPFGVNGCYNNALHGYEFGTNTSTDPKIGFKADSSSSSTLWGSGLNMTSDVGLVLYINDTSFGFLGWETITFDFIAPRTGYLGITLSSLDQDTIIMADQPTNFSKGSYLDSDGTPSLMDPALSLPNSSTYDTPNTLSPSFOLYQPKVAKYQTTNTNKLIEFYDTSATNTSLKLTITKNIKALIGGTASSOGNNNGGVSTINTITTTGNTSEGLKETSIOAETLKKFEDSKONKSEIIGDSTFTFKMDKLGAVYSTPLVNLINGGATSDSDTEKISFPGNOIDFNRLPTLPTLELPDPTNFYDYDYPPLVNLPSGFDKSTILKYSISVENQITGVRELFKDPOTQGLPVLNASTSTGPTVFQPPNOMADVPLPLTVPLVILVTLGLTIGIPMHRNKALQAGFDLSNRKVDVLTAKAVSFKEIINTGISNAFKLQATPTPKPKPTPKPKPKQ"

5402..8560

/note="114 kd protein"

/codon_start=1

/transl_table=4

/protein_id="AA25421.1"

/db_xref="GI:150160"

/translation="MKTMRKJOYKKAYWLLPLPLPLALNTFLVKDSKNATVATPEFTPTDTSKDLVSLAOLDSYQIADOTIHNTNLFLFKSRDVAIVKSSSSNNISPTFSOGKRSYVEETFNINIGIKMTMYKRYOLDVPNVSDMKQYAKNLIQLPULITYLNSSLAKKGTQREVLHLSGOANQMTSQRNHDNNPNPSSNASTGSEFNLYGNATRKLSWMPYEPIDGTQGGKGDSSGMSSTEENAKDAPVSGSGSSGFNLYGNATRKLSIGTLEDDQTPRNVITQLYASTSKLAVNNHLYVWGNFSLSMWVYSSIOEASNKPTAPANTNLMDGEDKOKQVFNQGLKETTSTINGNHSKSFQTPAVLISIDISDNGIIFSGFAGAGVGYDSSSSSSSTSDALAMTTSIDSKTQKXDLVMDGLNGLINGSFSTIDPESFYVYSSGNTNGTNGCTIKTAPVKKDQSTYKINSILNATPLNSYDGEICVFDAICLGNLNFKSNQERLPSRTDQIFVIGVISPENLRAKSSADSTGSDTKVMSNTQSRFLPVPYNIYSEGIIDAGFKRPERKASVTFSGLKAIDGFGANSLANFSVGLKAGIDPNPVMSGKRKANYGAVVLTGRGVRLNPNPNDLSLSTNTDNIAPISFS

misc.signal

/note="transcription termination signal"

BASE COUNT 2906 a 1704 c 1729 g 2421 t

ORIGIN 834 bp upstream of HincII site.

Query Match 3.9%; Score 124.6; DB 1; Length 8760;

Best local Similarity 63.5%; Pred. No. 2,1e-11;

Matches 190; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 2770 TTTGGCGCTTACCTTCATGGGTAGTGCCTACAGCAATTTGGATAGTATGTTCTT 2829

||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 8198 TTTGCAGCACTCCCTGCATGGGTATCCCTGTATCAGTATGTTCTTACGTGGATCTTG 8257

QY 2830 GCATTTATCATCATCTTACGATTAGTATCGGTATTCCTTTAAGAGCTCAAGAAATTA 2889

|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 8258 TTTATCTTGTAGCTTTAGACTTGGGATGGATCCCATGTACAGGGTAAAGAACTC 8317

QY 2890 CAAGACAAGGGGTTCAAAACAACATTCAAAAGTATGATACCTTGACCTGCTGTGGT 2949

||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 8318 CAAATGCACTCTGTTTATATGCTTTAAAAAGTTGATACCTCAACACTGCTGTGGT 8377

QY 2950 TCAGTTTACAGACAGATTTATTCACCACTGCTACGTTAGAAAAACCTGCTGTTTA 3009

|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 8378 AGTGCTACAAAAGATTTATTACCAACCTGGGTGTGAAAAACACCTAGTGCATGG 8437

QY 3010 GGTCGCTGAATCTTGTGTATAGAAACCTGCTGCTGCTGCTGAACCTGCTGCACG 3068

||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 8438 AAAGCTGATTAATCTTGTATTAACCACTGCTGCTTTTAAACCACCTGTTCAAC 8496

RESULT 11

939698

LOCUS Mycoplasma genitalium section 20 of 51 of the complete genome. 939698

DEFINITION U39698 L43967

ACCESSION U39698.1 GI:3844782

KEYWORDS

SOURCE

ORGANISM Mycoplasma genitalium

Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

REFERENCE 1 (bases 1 to 15787)

Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G.G., Kelley,J.M., Fritchman,J.L., Weidman,J.F., Small,K.V., Sandusky,M., Fuhmann,J.L., Nguyen,D.T., Uterback,T., Saudak,D.M., Phillips,C.A., Merrick,J.M., Tomb,J., Smith,H.O., Bult,C.F., Hu,P.-C., Lucier,T.S., Peterson,S.N., Smith,H.O. and Venter,J.C. The minimal gene complement of Mycoplasma genitalium

TITLE Science 270 (5235), 397-403 (1995)

JOURNAL 96026346

MEDLINE 7569993

PUBMED 2 (bases 1 to 15787)

Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G.G., Kelley,J.M., Fritchman,J.L., Weidman,J.F., Small,K.V., Sandusky,M., Fuhmann,J.L., Nguyen,D.T., Uterback,T.R., Saudak,D.M., Phillips,C.A., Merrick,J.M., Tomb,J.F., Dougerty,B.A., Bult,C.F., Hu,P.-C., Lucier,T.S., Peterson,S.N., Smith,H.O., Hutchinson,C.A. III and Venter,J.C. Direct Submissions

TITLE Submitted (29-OCT-1995) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

JOURNAL 3 (bases 1 to 15787)

Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G.G., Kelley,J.M., Fritchman,J.L., Weidman,J.F., Small,K.V., Sandusky,M., Fritchman,J.L., Weidman,J.F., Small,K.V., Sandusky,M.,

CDS

```

GSGDTRVKALKIEVKKSDSEEDNGQLOLEKNDLANAPIKRSEESGOSYQKADPCT
ALSSSGGSGNSPSPMRPMLATEIOIHDKLPKMSASILLIYDAEYANRPAIBDV
HLDRAMTANPRPSMTRKNNHLMKARVDLQOTGTFNRRHEPFDGQVAD
NEKTFGVDNSENTRKOFKOEKADSKSPYALPEEYAFAGIMLWFGQALLVFGNG
HVTSAHTAPLSIGFVRKYNATGDSAPYALPEEYAFAGIMLWFGQALLVFGNG
WFEYVPRMAVAGAVGRELVLACTITMGCTATVPRLYENPMORPISNLVAOGGILKE
DLQLETPYGANMRPDLIGAWSSSSSHNAPYFHNPNMORPISNLVAOGGILKE
DKNGDKAKYIYPRYSGMMAMOVYNNKLTDPQSLADVYENMAYOPNSIPALINP
SYLDQIDLYPVWIGNGYRGNHNGVDITAPOTSAGSSGISTNPSGSESLPTSN
IEVGLKANVQATIGSGSMTGAGSPRITLQANLQITODGALDQOATYTLPLPNL
TTSATGMOQSGSISACNPISIKODNISNIPYLVNRSQSDSKFQATDQKMSYDL
TPTADMPNALSFTNNKNAOKOLERIGLISIPYLVNRSQSDSKFQATDQKMSYDL
HSDQRLNLPAYEVNGLNLVSGTIVSFQLGWLVTFDPYKPRAGCTGLQITGLDASD
ATLITPGLAMTPQDVGNLVSGTIVSFQLGWLVTFDPYKPRAGCTGLQITGLDASD
ATORALINAPRPMAPRGSWNRGLRGSVNDLGVAMDQOSOSOSTTATRNALP
EHPNMLAPQVSYEASAKPNISGOSTNSPVLHVKRKYOTSDKLDLQITGLDASD
DPROVYTKIROSGFDHSTPOPOSLKTTTFVFGTSGSGLSVYSGGAGGSGSGQ
SGVDLSPRYKVSQMLVGOIPSTSDGNTNNLAVNTGNDVYGVYRLSESNAKKN
SDVDGLVTRPPLAELLIDGCTADGPOSKVFEKSPQIDENRLEFHPVYDLDPYMLV
YDOYIPLFIDIPASVNPKRKYLSPTDNMSIGLREFFKPPDQDOPNNVQVFN
NGDFLEPLTASSGQPTLSPNOMPDIYPLATIPVYIVLYLSTGLAIGIMHNK
KQALKAGFALSNOKVDLTKRAGSVFKEITNKGISQAPKRLQTSAKAPAPRPVP
PKRGAKEPVQPPKKA"
5943..9599
/note="orf-6; putative"
/codon_start=1
/transl_table=4
/product="p1 protein"
/protein_id="AA88326.1"
/db_xref="GI:1196658"
/translacion="MKSKLKRLRYLLEPLPLPLGLTSLANTYLLDHDHTLPYRPET
PINCGLIVRAAHLPKSYELVDKRVGDRLKALVLSALVRKFDOTTSDDSTNNON
ALSTDPQSKALNGSSQSSDSTGDSQSFASYVIFFAAPRANVADRTKILP
VKOSQSGQSGDSQSGSLKTTLODLEOPVYPPNGLARVNGVADDTVHESQ
ESSWNSQSGQGLKNNPKAVNGFKIDLSGAGSASISLQNGSNLSKILSASVSPSIS
SMKNSERTAEANDAPLVGMSGAGSASISLQNGSNLSKILSASVSPSIS
NOTLSLNPAPVQAVSQPAGGATAVSVNTAIDPTAFESKYLMTAOLKQMVY
PGLKMGNGNCTGVVASRODASTNLHLAGASOTGIGTSPPRATLTSORVAVGE
AGPLRAGNSSETDALPNVITOLYHTSTAOALVNGOIVWGSRPVSLWVWYGE
SGKATMAKTLKNGTOKOFENQGFQDSDSNLSKILSASVSPSIS
VYADHIVPAKAGAVGDMTDSASTYQALAMSTAGLSDGSKALVENTAID
GPKNCLITLIDTFAVTVPVSGMKGSONNEVQTVYKSDQKATKASLNASLN
STYGGDVATPVALGLENPKRNERLSPRDLVYIVNESELKASRENAQSTSDN
SNFVKWNTNASHILPYPYISANPAGNRBRKONGYKISTLESQATDGNESPL
NFTGLKAGVDPAPARSHKPNISAVLIRGCVRLMNPNDTKLSDYKRGYAVIT
SYTPGSAESAVDITLTKDYVIAESGLMTFTEPNGKRPYDCKOQKRRKYAVIT
VSRGTIERENEDANTLTSQAPALAVONGIASODDLGILPISDESAVITDQMT
GKVDYKNTNGLEKDDOLSENKVRKNDGLVPIYNEGIVDIMGWVDRANSVQANL
NPKYDEVINNDILQSFKEFTPAFANDRAMVEKESDITLVKPEITLIDNGTGE
DSKINGIPLINDPFRIFAGFALPSVIVPVSGSVGILLILILIGIGITPKVK
KLQDSSFDVPEKKVDTLITAVGSYVKIITQTSVIAKAPSKALANNAAPKAPVKA
PIAPRPVQPPKKA"
9617..9651
/note="putative"
/size="2560 c 2388 g 2194 t

```

terminator 2549 a 2560 c 2388 g 2194 t
 BASE COUNT 526 bp upstream of HindIII site.
 ORIGIN

Query Match 3.2%; Score 102.2; DB 1; Length 9691;
 Best Local Similarity 55.5%; Pred. No. 1.2e-07;
 Matches 197; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```

2737 AACATATTAGTACATCTCTATTAACACTTTGGCGCTTACCTTCACAGGTAGG 2796
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2734 AACATTGATTCCTCCGCGATTTGCTGCGCTTGTCCGTCGTCGGGCGAT 9293
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2797 CCTACAGCAATGGTACTAGTATCTGCACATATGATCATCTTAGCATAGCT 2856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2794 CCGGATCAGTCGCTTACGTCGTCGCTTCTTAATCTGCTGATCTTAGGCCCT 9353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2857 ATCGGTATTCCTTTAAGAGCTCAAGAAATTAACAGCAAGGCTTCAAAACACATTC 2916
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2854 ATTGGAATTCACATGATTAAGGTCCGCAAGCTTCAAGATTCAGCTTGTGATGTT 9413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY 2917 AAAAGGTGATACCTGACTGCTGCTGCTGCTGCTTACAGAGATATATACCCA 2976
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9414 AAAAGGTGATACCTGACTGCTGCTGCTGCTGCTTACAGAGATATATACCCA 9473
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2977 ACTGCTAGCTTAGGAAAAACCTGCTGCTTATAGTGTGCTGTAATCTGCTATAGAA 3036
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9474 ACGAGTGTGATCAAAAAAGCTCCTGATAGGCTGTAAAGCTGTAAATACCTGCTCTTAA 9533
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3037 CCGTCTGCTGCTGCTTAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3091
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9534 GCACAGTTAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 14
 AEO00002
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 1 (bases 1 to 16876)
 Hermann, R.
 Himmelreich, R., Hilbert, H., Plagens, H., Plick, E., Li, B. C. and
 Hermann, R.
 Complete sequence analysis of the genome of the bacterium
 Mycoplasma pneumoniae
 Nucleic Acids Res. 24 (22), 4420-4449 (1996)
 MEDLINE
 PUBMED
 8948633
 2 (bases 1 to 16876)
 Dandekar, T., Huynen, M., Regula, J. T., Ueberle, B., Zimmermann, C. O.,
 Andre, M. A., Doerks, T., Sanchez-Pulido, L., Snel, B., Suyama, M.,
 Yuan, Y. P., Hermann, R. and Bork, P.
 Re-annotating the Mycoplasma pneumoniae genome sequence: adding
 value, function and reading frames
 Nucleic Acids Res. 28 (17), 3278-3288 (2000)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 10954595
 3 (bases 1 to 16876)
 Himmelreich, R., Hilbert, H. and Li, B. C.
 Direct Submission
 Submitted (15-NOV-1996) Zentrum fuer Molekulare Biologie
 Heidelberg, University Heidelberg, 69120 Heidelberg, Germany
 4 (bases 1 to 16876)
 Suyama, M., Dandekar, T. and Hermann, R.
 Direct Submission
 Submitted (15-JUN-2000) Zentrum fuer Molekulare Biologie
 Heidelberg, University Heidelberg, 69120 Heidelberg, Germany
 On Nov 27, 2000 this sequence version replaced 91:1673651.
 This updated annotation replaces the old annotation from reference
 1. The old gene identifiers (MP numbers) according to the original
 publication by
 Himmelreich et al. (1996) are given as well as new gene numbering
 (MP numbers) from the origin of replication. Annotation comments
 and further update data are at
 http://www.bork.embl-heidelberg.de/Annot/MP/.
 Location/Qualifiers
 1..16876
 /organism="Mycoplasma pneumoniae"
 /mol_type="genomic DNA"
 /strain="M129"
 /db_xref="taxon:2104"
 /note="ATCC 29342"
 138..1112
 /gene="orf4"
 /note="synonym: E07_orf324"
 138..1112
 /gene="orf4"
 /note="MP1140(new), 015(Himmelreich et al., 1996)"
 /codon_start=1

FEATURES

source

gene

CDS

/transl_table=4
/product="DHH family phosphoesterases"
/protein_id="AAB95662.1"
/db_xref="GI:1673660"
/translation="MNSOVHRKGSIAEVAISIAQHDKIYFHHIRPDGCLGAHQGLA
RLIDQNFPRKOVCEVDPKPAHNEFMEVTFPEQITPELMQGLAVIVANXERIEC
RLIDLDNQKRAELRIDHEDNEDLNTNHFVDASYIAAEQVVDVLAQAKMLSPAA
TALYIGITDSNRELFYSNTSMRTLYLGSMLYRAQANIARIDHELNHTSLKIDFKOYV
FKNFOFONVIVYEVADKKROKKIKYVPLPCARNYLTIANIEQFIWLPFIEEGNNHRV
ERSNGINVREVALKIGGGGHHQASAVLKSKRDIIRVQDCOKQIAV"
1125..6008
/gene="p1"
/note="synonym: E07_orf1627"
1125..6008
/gene="p1"
/note="MPN141(new), 014(Himmelreich et al., 1996)"
/transl_table=1
/codon_start=1
/product="ADP1_MyCPN adhesin p1"
/protein_id="AAB95661.1"
/db_xref="GI:1673659"
/translation="MHQTKTALSSTWILLITATASLTAGTGVGHFTSTTTLRQ
QSYTRDEVALHNTAINRPLTPWYRNNTSESLDPTGENPGAMALVNDNSKGLTA
GSGSQQTTPDPTREATAALASTFTALRYDLAARALYDLFSLKLNROTPTROTQIT
FNFEGFGLSGAAPQOMNEKVKVPEVADSPNPRFALVPRSVVYEOLQRIIG
LPOQRTSGONTSTTGAMFGKYNAAEDTAKNEKLOGEATSGSTSGSGSTORG
GSGDTRKVAIKTEVKKKSDSDNCOLEKNDLANAPIKRSESGSOVAKADDCGT
ALSSSGSGSNPGSPTPWRPLATBOTIKHDKPKMSASTILITADYARKTATIDVYD
HIDPKAMTANYPSPMTRPKWNHGLMDMKARDVLLQTGFENRHRPEWFDGQIYAD
NERTQDVENSENTEKQFQKEADSDAPIALPFEAYFANIGLWFGQALVFGQV
HYKSAHTAPLSIGVRRVFNATGTSATYGMVYALFSGMVKQTDGLKQDFPNR
WFEYVPRMAVAGAKFVGRRELYLAGTITMGPTAVPRLLDELSNLTNLAOGGGLRE
DLQLETPYGANRPLPIGAMSSSSSSSSSSNAPYFNNNDODRPIONVDAIRKWE
DKRKDQAKYIYPRYSYGAMQVYWNKSLTQPIASDFVNANAPNSLPLAIVG
ELLALPDKYKKEKNEFANETERNOKUTVAPTOGTWNSHPSTLSRSTFNLIN
SVLDQVLDVYVPIGNGYRYGNHNGVDITAPDTSGSSSGISNTSGSRPLPTSN
IGVGLANVOATLGSGQWTITGSPRRITDQANLQMTGAGMRNDKASSQSDENHTK
FTSATGMDQOGSGTSAGNDPSLKODNISKSGSLTODGNAIDQOEATYTNLPNL
TPTADPNALSTFNKNNAORALFLRGLIGSTPIVLYNNGSDNKRFOATDOKKSYDL
HSDPTKLNLPAYGEVGLNLPALVEYFGRTRAGSGSNTSSPGIGFKTPEPNDSK
ATLITPGLAMTPDQVGNLVVSGTTSVFOGLGVITFTDVKPRAGVLGQLDTGLDSD
ATQRALIMAPRPAAFRGSMVNRGLGREVMDKGVADAOQSDSGSTTATARNLP
EHNALAFQVSVVEASAYKPNNTSSQGTQSNSSPYLIVKPKVTOGSKDLDDKLNL
DPNOVKTRLOSFGDSTHSTOPOLKTTTPVGTSSGNLSVLSGGGAGSGSGSG
SGVDLSPEVKVSGMLVGLPSTSDGNTSSNAPNTNNGNDVYVGRISSESAAMN
DDVDGIVRTPPLALDGEQOTADTPGQSVKFKSPDIDENRLLTHPYTDLFDVYMLY
YDQYIFLEIDIPASVNPKNVRLKVLVSDNEOSGLRLEFKPDQDTQPNNNQVFN
NGDFLPLTASSQGPOTLFSFPMQMPDYVLPATIVPIVIVYLSVLGALIGIPMKN
KQALKAGFALSNQKVDVLTKAVGVSVEKEIINRTGISGAPRKLKOTSAKGAERPRPV
PKGAKKPPVQPKKPA"
1125..2671
/rpt_family="REPMP4"
/rpt_family="REPMP4"
3171..4971
/rpt_family="REPMP2/3"
6014..9670
/gene="E07_orf1218"
/note="synonym: orf6"
6014..9670
/gene="orf6"
6014..9670
/gene="orf6"
6014..9670
/note="MPN142(new), 013(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="involved in cytaadherence"
/protein_id="AAB95660.1"
/db_xref="GI:1673658"
/translation="MKSRLIKRYLLEPLPLPTGTSIANTYLLLODHNTLPTPTPT
PLNGLDIVRAALHPSYELVDMKRVGDRTKIVLASALVRKEDPTSSDOSNTON
ALSEPQESOKALHNSQSGSDTSNDSODFAVYLIFKAAPRATWYFEKIKTALPY
VKQESGSGDQSGNGKSLKTLQDLLEDPVTPYTPNGLARVNGVADQDTVHFGSGQ
ESSWNSQRSOKKNNPGRKAVTGFKLDKRAVRKLNESVPEVPELSTDEGKGDES
SMKNSERTTAENDAPLVGMVSGAAGSASSLSQNGSNSGLKSLRSAPVSPSSTS

NOTLSLNPAPVGPQAVVSPQAGATAAVSNRTASDTATFSKYLNTAQALHQMGVIV
PGLEKMGNGNGVAVSRQDATSTNLPHAAAGSQTGILVWGSGRPSVLTWTSORVAVV
AGPLRAGNSSENTDALPNVITQLYHTSTADALDLSNGSSKSNLSKALQGLTPVLYAGLD
SGRATWMAKTELMLWGTDKOKQFVENOLGRKADLNSQSNKSNLSKALQGLTPVLYAGLD
VVAHDILVEAFAPAGAVGYDMTIDSSASTYNQALAMSTYTAGLSDGGYKALVENTALN
GPINGLFTLDFEFAVYTPVSGMGGSQNNDEQVTPVPSDQKATKAKISLNASLN
SYGDQVTEVADLGNFNKLINEERLSPRDEQDLVYGINESLKSARNAQSTSDN
SNKVCVMTASHYLTPVYVYSANPEACNRRRABQNRGVKISTLESQANDGFANSL
NCTGGLKACVDAPPAVRGKRPNSAVLLYRGVYVRNENPDQDLDDSTDKNEPISF
STPRTGSASVADLTLLKQVITYIAESGLWFYTFDNEKFTYDQKQVKNRKYAVIT
VSRGTGEEFNEDANTTLLSQAPALAVQNGIASSQDQDTGLPLSDFSAVITKQDTWT
GKVDYKNTNGNGLEFKDDQLSENVKRRNDGIVPYNDEIVDINGRVDFAVANSYQANL
TKQTVDEVINNPDILOSFFKFPFAPNORAMLYGKTSPTDITVYKIEYLDGNFYGE
DSKIAGIPLNIDEPSPRIIFAGPALPSVWIPVSGSSVGTILLIILIGLIGIPMYKVR
KLQDSSFVDVPEFKKVPDLITRANGSVYKKIITQTSVIRKAPSALKANNAPKAPVRKA
PTRAHPVQPKKPA"
6014..8714
/rpt_family="REPMP5"
9787..10314
/gene="E07_orf175"
9787..10314
/gene="E07_orf175"
/note="MPN143(new), 012(Himmelreich et al., 1996)"
/transl_table=1
/codon_start=1
/product="hypothetical protein"
/protein_id="AAB95659.1"
/db_xref="GI:1673657"
/translation="MSHKDFNGLOAPOLSSSPVAKKQSHKRLRALHARYLNHSS
KRLKHALEHEDNOVLLREKSGPNQDMLSKQPGYNKSLKLNKLSIGVSKESPK
KRPPPTFGKSTPTPEEKALQOMQOSKRFPEHNMHSPLNEVANNPMIQRKQKQA
KRAANRORTYKRO"
10457..11977
/rpt_family="REPMP2/3"
10888..12129
/gene="E07_orf413"
10888..12129
/gene="E07_orf413"
/note="MPN144(new), 011(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="hypothetical protein, see: MPN141"
/protein_id="AAB95658.1"
/db_xref="GI:1673656"
/translation="MGQOGSGTSAGNPSISKODKTSKSGSLITDQNGATGQOEATN
YTNLPNLPPTADPNVALSTFNKNHRAOLFRLGLGSPVLYVNSGSDSNKFOATN
QKWSYTDLOSQDKLNLPAVGEVGNLNPALVETFGNTRAGSGSNTSSPGIGFKI
PEGNDSKATLITPGLAMTPDQVGNLVVSGTSLSPQGLWVFTFDIIPRAGYALIG
LSGLDASDSQDRELINAKRMAFRGSMVNRGLGREVMDKGVADAOQSDSGSTTATARNLP
EASGSALEPPLNALFOVSVVEASAYSSSTSSSGSGSSNTSPYLIHFKKVEESTTO
LDQGLKLNLDPRNOVTRKLRQSGTSDHSTQPOSILKTTTPVFGTSSGNIGVLSGGGAGG
GSSGSGSGVDLSPEVRSQH"
12481..13020
/gene="E07_orf179"
12481..13020
/gene="E07_orf179"
/note="MPN145(new), 010(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="hypothetical protein, see: MPN013"
/protein_id="AAB95657.1"
/db_xref="GI:1673655"
/translation="MSYSPSLKELITAILOKYTSKNYQSNCKRPPDKIELSLNGVFE
IKVTPKRTIVTHKQDLQKDFKQDLMELHPTFKTDLAKSEARINOKTALLOI
VWQEGQIKVHGEOINKLTQVAVEKGEKEIEAQOQIOKQVNETINFEVESIGSIHKRLD
SMEGRLDSMENRDLKLESK"
13338..14135
/gene="E07_orf265"
13338..14135
/gene="E07_orf265"
/gene="E07_orf265"
Query Match 3.24; Score 102.2; DB 1; Length 16876;
Best Local Similarity 55.5%; Pred. No. 1.1e-07;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 06:06:51 ; Search time 798 seconds
(without alignments)
10787.604 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189
Sequence: 1 atgaatattctcaaaaact.....caaccgtcccaagaataa 3189

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq-19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3144.2	98.6	3189	18	DNA encoding antiq
2	3143.4	98.6	10651	22	Vector pCDNA3-Fast
3	124.6	3.9	580073	18	Mycoplasma genital
4	78.4	2.5	4182	16	DNA encoding cyta
5	64.6	2.0	3435	16	Mycoplasma plium a
6	62.6	2.0	4985	24	Anopheles gambiæ
7	60.8	1.9	8771	24	Human immune syste
8	60.2	1.9	3489	21	Kaposi's sarcoma-a

C	9	60.2	1.9	3489	22	AAE82901	Nucleotide sequenc
C	10	60.2	1.9	3489	24	ABA93487	Kaposi's sarcoma-a
C	11	60.2	1.9	32207	20	AAV73805	KSHV LUR DNA (nucl
C	12	60.2	1.9	137507	19	AAV19941	KSHV long unique c
C	13	59	1.9	535	24	ABO42292	Oligonucleotide fo
C	14	59	1.9	535	24	ABO42293	Oligonucleotide fo
C	15	58.4	1.8	640681	24	ABA92787	Buchnera sp. genom
C	16	58	1.8	50000	24	ABA55644	AmphV genome fragm
C	17	55.8	1.7	969	24	ABO39490	Oligonucleotide fo
C	18	55.8	1.7	969	24	ABO39491	Oligonucleotide fo
C	19	54.8	1.7	3837	21	AAA70211	Plasmodium falcipla
C	20	54.6	1.7	14041	22	AAH48024	Internal control B
C	21	53.8	1.7	3582	21	AAA70241	Plasmodium falcipla
C	22	53.4	1.7	1159	21	AAA59240	An EcoRI fragment
C	23	53.2	1.7	987	18	AAE90555	Clostridium perfr
C	24	53.2	1.7	1037	21	AAA59242	Clostridium perfr
C	25	53.2	1.7	1472	21	AAA59241	Exons E, C and A o
C	26	53.2	1.7	8056	25	AAE10246	Exons D, C, B and
C	27	53	1.7	987	18	AAE90554	Haematopoietic cel
C	28	53	1.7	13784	24	ABR40062	Clostridium perfr
C	29	52.4	1.6	3057	20	AAE99555	Human chemically p
C	30	52	1.6	972	20	AAE61754	Nucleic acid seque
C	31	52	1.6	1029	20	AAE61753	B. burgdorferi ant
C	32	52	1.6	5120	22	AAE84677	B. burgdorferi ant
C	33	52	1.6	910715	20	AAE20248	DNA sequence of hu
C	34	51.8	1.6	8056	25	AAE10100	Borrelia burgdorfe
C	35	51.6	1.6	2451	22	AAE07045	Haematopoietic cel
C	36	50.6	1.6	901	23	AB108561	Drosophila melanog
C	37	50.6	1.6	6087	23	AB108560	Drosophila melanog
C	38	50.2	1.6	2010	24	AAE18114	Streptococcus dysg
C	39	50	1.6	2355	24	ABW71061	Streptococcus poly
C	40	50	1.6	2448	24	ABW65054	Streptococcus poly
C	41	50	1.6	50000	24	AB155643	AmphV genome fragm
C	42	49.8	1.6	215561	24	ABN71527	Streptococcus poly
C	43	49.8	1.6	486	22	AAE75507	AmphV genome fragm
C	44	49.4	1.5	1887	20	AAE9539	Polyglutamine trac
C	45	49.2	1.5	40324	24	ABO67150	Nucleic acid seque
							Human angiogenesis

ALIGNMENTS

RESULT 1	AAE75087	AAE75087 standard; DNA; 3189 BP.
ID	AAE75087	
AC	AAE75087	
XX		
DT	27-FEB-1998	(first entry)
XX		
DE	DNA encoding antigenic protein derived from Mycoplasma gallisepticum.	
XX		
KW	Mycoplasma gallisepticum antigen; epitope: monoclonal antibody: bird;	
KW	Mycoplasma infection; viral vaccine; avipoxvirus; herpesvirus; ss.	
XX		
OS	Mycoplasma gallisepticum.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..3189
FT		/tag= a
FT		/product= Mycoplasma-antigen
FT	CDS	1..3189
FT		/tag= b
FT		/codon= (seq: tga, aa: trp)
XX		
PN	WO9724370-A1.	
PD	10-JUL-1997.	
XX		
PF	27-DEC-1996.	96MO-JP03863.
XX		
PR	28-DEC-1995.	95JP-0352754.
XX		

PA (JAPG) NIPPON ZEON KK.
 XX Fujisawa A, Yoshida S;
 PI
 XX MPI: 1997-363621/33.
 DR P-PSDB; AAW22721.
 XX
 PT Antigenic protein derived from Mycoplasma gallisepticum - useful in
 vaccines against, and for diagnosis of mycoplasma infection in birds
 XX
 PS Claim 4; Pages 24-31; 45pp; Japanese.
 XX
 CC This sequence represents DNA encoding an antigenic protein of Mycoplasma
 CC gallisepticum having an epitope specifically recognised by the monoclonal
 CC antibody (Mab) 35A6. This antigenic protein inhibits the metabolism of
 CC bird infecting mycoplasmas. The antigenic protein and the DNA molecule can
 CC be used in the preparation of component and viral vaccines against
 CC mycoplasma infection in birds. Mycoplasma infection can be diagnosed by
 CC hybridising DNA in a sample with a DNA probe corresponding to 10 or more
 CC (preferably 15 to 40) bases of the antigenic protein gene.
 CC
 XX
 SQ Sequence 3189 BP; 1130 A; 572 C; 522 G; 965 T; 0 other;
 Query Match 98.6%; Score 3144.2; DB 18; Length 3189;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 3161; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 1 ATGAATATTTCTAAAAAAGTTATACATGTAGAGTGATAGCTGTATTTGA 60
 DB 1 ATGAATATTTCTAAAAAAGTTATACATGTAGAGTGATAGCTGTATTTGA 60
 OY 61 GCTCTGGTCTGCAGACCTTTGGCTTTAGCAATCAGATTAAGCTTAACACGGCA 120
 DB 61 GCTCTGGTCTGCAGACCTTTGGCTTTAGCAATCAGATTAAGCTTAACACGGCA 120
 OY 121 TTAGTTATCAGCAAGACGCTGATGATGATCTGTTAGACTGAGCTTTGGACA 180
 DB 121 TTAGTTATCAGCAAGACGCTGATGATGATCTGTTAGACTGAGCTTTGGACA 180
 OY 181 AATGCTGCTGTTCCATACAGTCTTAGAGATGTTATGATTAACCTTTAAGACGCT 240
 DB 181 AATGCTGCTGTTCCATACAGTCTTAGAGATGTTATGATTAACCTTTAAGACGCT 240
 OY 241 AATGCAACATTTATTAATTTAGATGTTTCTAACAACATTAATGTTAGATCTAAGT 300
 DB 241 AATGCAACATTTATTAATTTAGATGTTTCTAACAACATTAATGTTAGATCTAAGT 300
 OY 301 GATGATTTGCTGATACAAAGTAAACAAATAGTTTCAATGATGATGATGAT 360
 DB 301 GATGATTTGCTGATACAAAGTAAACAAATAGTTTCAATGATGATGATGATGAT 360
 OY 361 AGATTTGATCAAGCAACAGACATTAATGCTGTTGTTGTTAAGATGAGACCTAAC 420
 DB 361 AGATTTGATCAAGCAACAGACATTAATGCTGTTGTTGTTAAGATGAGACCTAAC 420
 OY 421 GTTCAATTTAAAAAGATTATACCTAATAGATGTTATAGAAACAACATTTCT 480
 DB 421 GTTCAATTTAAAAAGATTATACCTAATAGATGTTATAGAAACAACATTTCT 480
 OY 481 AAGTTGTAATGCTGTTGTTGATTAATCCAGCTCAGCTGATTAAGATTACTGATGAGG 540
 DB 481 AAGTTGTAATGCTGTTGTTGATTAATCCAGCTCAGCTGATTAAGATTACTGATGAGG 540
 OY 541 ACTAATTTAATTTTACAAACCAACCTCAAGTGAATTTGTAATGACTTCAATTTAGAT 600
 DB 541 ACTAATTTAATTTTACAAACCAACCTCAAGTGAATTTGTAATGACTTCAATTTAGAT 600
 OY 601 GCGCAATCTTACCTAAAGATTTACACCCAGATGATTAATCACTTACATTCAGAAAG 660
 DB 601 GCGCAATCTTACCTAAAGATTTACACCCAGATGATTAATCACTTACATTCAGAAAG 660
 OY 661 ATCTTACCAATGACGTCACACATGACGTTGTTCTTGGCCAGTAGTAGAGTTAGTGA 720
 DB 661 ATCTTACCAATGACGTCACACATGACGTTGTTCTTGGCCAGTAGTAGAGTTAGTGA 720

DB 661 ATCTTACCAATGACGTCACACATGACGTTGTTCTTGGCCAGTAGTAGAGTTAGTGA 720
 OY 721 ACAATGCTGATGATGGATGTTGATTTGGGAATGTCAAATACATACATGCT 780
 DB 721 ACAATGCTGATGATGGATGTTGATTTGGGAATGTCAAATACATACATGCT 780
 OY 781 ATTGCTCAATTAACACACTGATGATTAACAAATCTTCAACTTTTAATCAGAGCA 840
 DB 781 ATTGCTCAATTAACACACTGATGATTAACAAATCTTCAACTTTTAATCAGAGCA 840
 OY 841 ATGCTGATGCAACCAATAGATACGATTTCAATGTAATGTCAGATGATTAATAACA 900
 DB 841 ATGCTGATGCAACCAATAGATACGATTTCAATGTAATGTCAGATGATTAATAACA 900
 OY 901 TCTTCCAAATAGATGAATAATTTGTTATCCAGATGAGCTGTTGAGAGATTA 960
 DB 901 TCTTCCAAATAGATGAATAATTTGTTATCCAGATGAGCTGTTGAGAGATTA 960
 OY 961 AATATTACAGATTTAGCTACGAGTTGCGCAACGCAAGCAAGATTTGATTTCTTGAC 1020
 DB 961 AATATTACAGATTTAGCTACGAGTTGCGCAACGCAAGCAAGATTTGATTTCTTGAC 1020
 OY 1021 ATACCGGAGCTCCCAAGTTACTTTAAAGAGATTCAGTTAAGATTTTCAAGCTA 1080
 DB 1021 ATACCGGAGCTCCCAAGTTACTTTAAAGAGATTCAGTTAAGATTTTCAAGCTA 1080
 OY 1081 TACTTAACCTAGATTAATCTTATCATTCATTCATTCATTCATTCATTCATTCATTC 1140
 DB 1081 TACTTAACCTAGATTAATCTTATCATTCATTCATTCATTCATTCATTCATTCATTC 1140
 OY 1141 TCTGAATTTACCATCATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 DB 1141 TCTGAATTTACCATCATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 OY 1201 TTGAATCAAGTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1201 TTGAATCAAGTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 OY 1261 AAGGAAACAGCAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1261 AAGGAAACAGCAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 OY 1321 ACTACTAACCTTCTCAACAGTTCTTAATCTTCAATCTTCAATCTTCAATCTTCAATCT 1380
 DB 1321 ACTACTAACCTTCTCAACAGTTCTTAATCTTCAATCTTCAATCTTCAATCTTCAATCT 1380
 OY 1381 ATTGATTAACCAACCTTCTGCAACCAATAGATGATGATGATGATGATGATGATGAT 1440
 DB 1381 ATTGATTAACCAACCTTCTGCAACCAATAGATGATGATGATGATGATGATGATGAT 1440
 OY 1441 GTTATTTGAAGCAAGATTAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1441 GTTATTTGAAGCAAGATTAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1500
 OY 1501 ACTAATGCAAGAACTTTATGCGAAGCAATGTTGTTGTTGTTTACTTCAACAGT 1560
 DB 1501 ACTAATGCAAGAACTTTATGCGAAGCAATGTTGTTGTTGTTTACTTCAACAGT 1560
 OY 1561 TCAGAGTGTGTTTAAAGCTTCTTAATACGCTGATCAAGCTCAACTGAAACTTCA 1620
 DB 1561 TCAGAGTGTGTTTAAAGCTTCTTAATACGCTGATCAAGCTCAACTGAAACTTCA 1620
 OY 1621 CCTTCTTATACCTATTTGTTTATGATTAACCAACCACTAGAAACGAACTTCTGGA 1680
 DB 1621 CCTTCTTATACCTATTTGTTTATGATTAACCAACCACTAGAAACGAACTTCTGGA 1680
 OY 1681 TACGGAACATATACCTTTTAAACCAACAGCCCTTACGAGCTTATGATTTCTCCAAAGATA 1740
 DB 1681 TACGGAACATATACCTTTTAAACCAACAGCCCTTACGAGCTTATGATTTCTCCAAAGATA 1740
 OY 1741 GGTACTGAACCAATCAATTTAGAAAGACTTCAATTAACCTACCTGTTATGGTGATAT 1800
 DB 1741 GGTACTGAACCAATCAATTTAGAAAGACTTCAATTAACCTACCTGTTATGGTGATAT 1800

```

OY 1801 CTAAGTGAAGAGTGTGAGTATTCCTCTAATACCTCATATATAAGACACAGGTGAC 1860
    |||||||
DB 1801 CTAAGTGAAGAGTGTGAGTATTCCTCTAATACCTCATATATAAGACACAGGTGAC 1860
OY 1861 ACACGAGAAGGCCGAGAGCATCTTCCAAATCTGGCTATTCGTATATACCTATGAGTACAT 1920
    |||||||
DB 1861 ACACGAGAAGGCCGAGAGCATCTTCCAAATCTGGCTATTCGTATATACCTATGAGTACAT 1920
OY 1921 CAATACGTTTATAGATTTGATGGAATTAAGTAATTAATGTTGGGTTAAAGCATCA 1980
    |||||||
DB 1921 CAATACGTTTATAGATTTGATGGAATTAAGTAATTAATGTTGGGTTAAAGCATCA 1980
OY 1981 AGCTTCTTAACTCAAAATAGACCAATCCAAAGGCTCTGAAATGTTGCTGCAACACA 2040
    |||||||
DB 1981 AGCTTCTTAACTCAAAATAGACCAATCCAAAGGCTCTGAAATGTTGCTGCAACACA 2040
OY 2041 TACTTAAGTCAACAATTTGATTTAGCTAGACATCTGTATACCAACCAACACCATTC 2100
    |||||||
DB 2041 TACTTAAGTCAACAATTTGATTTAGCTAGACATCTGTATACCAACCAACACCATTC 2100
OY 2101 GGAACAACGTCACCAAGTTATTTGATATACCTGCTGATGATGATGATGATGATGAT 2160
    |||||||
DB 2101 GGAACAACGTCACCAAGTTATTTGATATACCTGCTGATGATGATGATGATGATGAT 2160
OY 2161 ATTAGAACAATCTTCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2220
    |||||||
DB 2161 ATTAGAACAATCTTCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2220
OY 2221 AAATCTAGTGTATTAATTAAGATTAGCTGATGATGATGATGATGATGATGATGATGAT 2280
    |||||||
DB 2221 AAATCTAGTGTATTAATTAAGATTAGCTGATGATGATGATGATGATGATGATGATGAT 2280
OY 2281 TTCAGTCCACAAAGTTATTAATTAAGATTAGCTGATGATGATGATGATGATGATGATGAT 2340
    |||||||
DB 2281 TTCAGTCCACAAAGTTATTAATTAAGATTAGCTGATGATGATGATGATGATGATGATGAT 2340
OY 2341 AATTCATCTATACAGTAATATGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 2400
    |||||||
DB 2341 AATTCATCTATACAGTAATATGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 2400
OY 2401 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
    |||||||
DB 2401 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
OY 2461 GAAGGTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
    |||||||
DB 2461 GAAGGTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
OY 2521 GAGTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2580
    |||||||
DB 2521 GAGTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2580
OY 2581 ACCTTCTTAACAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
    |||||||
DB 2581 ACCTTCTTAACAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
OY 2641 GTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
    |||||||
DB 2641 GTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
OY 2701 TACTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
    |||||||
DB 2701 TACTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
OY 2761 TATTAACACTTTGGGCTTACCTTCATGATGATGATGATGATGATGATGATGATGATGAT 2820
    |||||||
DB 2761 TATTAACACTTTGGGCTTACCTTCATGATGATGATGATGATGATGATGATGATGATGAT 2820
OY 2821 GGTATTTCTGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880
    |||||||
DB 2821 GGTATTTCTGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880

```

```

OY 2881 AGAAATTAACAAGAGGTTCAAAACCAATTCACAAAAGTTGATACCTTGACTGCT 2940
    |||||||
DB 2881 AGAAATTAACAAGAGGTTCAAAACCAATTCACAAAAGTTGATACCTTGACTGCT 2940
OY 2941 GCTGTGTTGCTTACAGTAAGATTAATTAACCAAACTGCTTAACGTTAAGAAAAAACC 3000
    |||||||
DB 2941 GCTGTGTTGCTTACAGTAAGATTAATTAACCAAACTGCTTAACGTTAAGAAAAAACC 3000
OY 3001 GCTGTGTTGCTTACAGTAAGATTAATTAACCAAACTGCTTAACGTTAAGAAAAAACC 3060
    |||||||
DB 3001 GCTGTGTTGCTTACAGTAAGATTAATTAACCAAACTGCTTAACGTTAAGAAAAAACC 3060
OY 3061 GCTCCAGCTTAACCAATCTGACCAAAAGCTGACCTACAGTAACCAACTGAGCTTAA 3120
    |||||||
DB 3061 GCTCCAGCTTAACCAATCTGACCAAAAGCTGACCTACAGTAACCAACTGAGCTTAA 3120
OY 3121 TCTGTGTCGCTTACAAACCAACTGCTCTAAGCCAGCTGCTCCAAACCAACCGCTCCC 3180
    |||||||
DB 3121 TCTGTGTCGCTTACAAACCAACTGCTCTAAGCCAGCTGCTCCAAACCAACCGCTCCC 3180
OY 3181 AAAGCAATTA 3189
    |||||||
DB 3181 AAACAAATTA 3189

```

RESULT 2
AAF88025
ID AAF88025 standard; DNA; 10651 BP.

AAF88025;
16-JUL-2001 (first entry)

DE Vector pcdNA3-FasL-IRES-crma DNA.

XX KW Vector; pcdNA3-FasL-IRES-crma; autoimmune disease; apoptosis; crma;
KW suicide enzyme; anti-inflammatory; immunomodulatory; rheumatic;
KW arthritic; dermatological; gene therapy; rheumatoid arthritis;
KW systemic lupus erythematosus; Sjogren-Syndrome; polymyositis;
KW dermatomyositis; polymyalgia; rheumatism; arteritis temporalis;
KW Crohn's disease; Behcet's disease; colitis ulcerosa; thyroiditis;
KW autoimmune hepatitis; diabetes mellitus (type I); psoriasis; dermatitis;
KW pemphigus vulgaris; multiple sclerosis; myasthenia gravis;
KW viral infection; brain inflammation; transplant rejection; ds.

OS Synthetic.
XX PN WO200127254-A2.
XX PD 19-APR-2001.
XX PF 12-OCT-2000; 2000MO-DE03608.
XX PR 12-OCT-1999; 99DE-1048983.
XX PA (SCHW/) SCHWARZMANN F.
XX PI Schwarzmenn F.
XX DR WPI; 2001-282021/29.
XX PT New gene transfer vector useful in the prevention and treatment of
XX autoimmune and chronic inflammatory disease -
PS Claim 9; Page 74-79; 82pp; German.

CC This invention describes a novel gene transfer vector (NI) which
CC comprises (1) a first nucleic acid sequence (I) encoding one or
CC more ligands that trigger apoptosis; (2) a second nucleic acid sequence
CC (II) encoding one or more antigens; and optionally (3) a third nucleic
CC acid sequence (III) encoding one or more anti-apoptosis molecules; and
CC optionally (4) a fourth nucleic acid sequence (IV) encoding one or more
CC suicide enzymes. The products of the invention have anti-inflammatory,

Db 1806 CTAAGTGAAGAGTGTGAGAAAGTTCTCTAATTAATCTCATATATATTAAGACACAGGTGAC 1865
 QY 1861 ACACGAGAAGCCGAGACATCTTCCATCTGCGTATTCGTATGATATATAGTACATT 1920
 Db 1866 ACACGAGAAGCCGAGACATCTTCCATCTGCGTATTCGTATGATATATAGTACATT 1925
 QY 1921 CAATCAGTTTTAGATTTTGTGAGAAATAGAAATTAATTAATGTTGGGTTAAAGCATCA 1980
 Db 1926 CAATCAGTTTTAGATTTTGTGAGAAATAGAAATTAATTAATGTTGGGTTAAAGCATCA 1985
 QY 1981 AGCTCTTAATCTCAATAGACCAATCCAAAGGCTCAGAAATATGTTGCTGCAACACCA 2040
 Db 1986 AGCTCTTAATCTCAATAGACCAATCCAAAGGCTCAGAAATATGTTGCTGCAACACCA 2045
 QY 2041 TACTTAAGATCAACAATTTGATTTAGCTAGAACATCTGATTTACCAACCAACCATTC 2100
 Db 2046 TACTTAAGATCAACAATTTGATTTAGCTAGAACATCTGATTTACCAACCAACCATTC 2105
 QY 2101 GGACACACTCACCAAGTTATTTCACTATCACCTGATCACTTCATCAATTAAGANT 2160
 Db 2106 GGACACACTCACCAAGTTATTTCACTATCACCTGATCACTTCATCAATTAAGANT 2165
 QY 2161 ATTAAACAATCTCCGCTGATCAAGTATGCTATCTCTTATTCACAATGAAATATAT 2220
 Db 2166 ATTAAACAATCTCCGCTGATCAAGTATGCTATCTCTTATTCACAATGAAATATAT 2225
 QY 2221 AAATCTAGTGTATTATACATTAATAGATTAAGTACTGACTCAAGTACCTGATCAAGCTCA 2280
 Db 2226 AAATCTAGTGTATTATACATTAATAGTACTGACTCAAGTACCTGATCAAGCTCA 2285
 QY 2281 TTCACTCCAAACAAGTTTAATTTAGACGTAAATGAATTTGTTGTAATCTTACCTTTATTAGAC 2340
 Db 2286 TTCACTCCAAACAAGTTTAATTTAGACGTAAATGAATTTGTTGTAATCTTACCTTTATTAGAC 2345
 QY 2341 AATTCATCTTAAGAGTAAATGCTGCTGATGTAATGTTGTAATGTTGTAATGTTGTAATGTTGTA 2400
 Db 2346 AATTCATCTTAAGAGTAAATGCTGCTGATGTAATGTTGTAATGTTGTAATGTTGTAATGTTGTA 2405
 QY 2401 TCTCTGATCATATATCTGCTGATTAATATCAATTAATCAAGTATCTGATATGCTTTT 2460
 Db 2406 TCTCTGATCATATATCTGCTGATTAATATCAATTAATCAAGTATCTGATATGCTTTT 2465
 QY 2461 GAAGCTTCTGCTGATTAATATCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
 Db 2466 GAAGCTTCTGCTGATTAATATCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2525
 QY 2521 GAGTACTTAATCAAAATGGGTTCACTACTCAAGTCCGCTAGAACTCTGCTGCTGCTGCTGCTGCTGCT 2580
 Db 2526 GAGTACTTAATCAAAATGGGTTCACTACTCAAGTCCGCTAGAACTCTGCTGCTGCTGCTGCTGCTGCT 2585
 QY 2581 AGCTCTTAATCAAAATGGGTTCACTACTCAAGTCCGCTAGAACTCTGCTGCTGCTGCTGCTGCTGCT 2640
 Db 2586 AGCTCTTAATCAAAATGGGTTCACTACTCAAGTCCGCTAGAACTCTGCTGCTGCTGCTGCTGCTGCT 2645
 QY 2641 GTGGTTGATCTGATGTAATTAACAAACCAAACTCTAATAAGTTGATGATCA 2700
 Db 2646 GTGGTTGATCTGATGTAATTAACAAACCAAACTCTAATAAGTTGATGATCA 2705
 QY 2701 TACTTAAGATGTAATATGATGCTAATTAAGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
 Db 2706 TACTTAAGATGTAATATGATGCTAATTAAGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2765
 QY 2761 TATTAACAATTTGGGCTTACCTCATGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 Db 2766 TATTAACAATTTGGGCTTACCTCATGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2825
 QY 2821 GGTATTTCTGCAATTAATGATCACTTATGATTAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
 Db 2826 GGTATTTCTGCAATTAATGATCACTTATGATTAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2885
 QY 2881 AGAAAAATTAACAAGAGGTTCAAAACCAATTTAAAAAGTTGATACCTTGACTGCT 2940
 Db 2886 AGAAAAATTAACAAGAGGTTCAAAACCAATTTAAAAAGTTGATACCTTGACTGCT 2945

QY 2941 GCTGTTGGTTTCAGTTTACAGAAGATTATTACCAAACTGCTAACGTTAAGAAAAAAGCT 3000
 Db 2946 GCTGTTGGTTTCAGTTTACAGAAGATTATTACCAAACTGCTAACGTTAAGAAAAAAGCT 3005
 QY 3001 GCTGCTTTAGTGTGCT 3060
 Db 3006 GCTGCTTTAGTGTGCT 3065
 QY 3061 GCTGCTTTAGTGTGCT 3120
 Db 3066 GCTGCTTTAGTGTGCT 3125
 QY 3121 TCTGCT 3180
 Db 3126 TCTGCT 3185
 QY 3181 AAGAAATA 3189
 Db 3186 AAGAAATA 3194
 RESULT 3
 AAT58840
 ID AAT58840 standard; DNA; 580073 BP.
 AC AAT58840;
 XX
 DT 27-MAR-1997 (first entry)
 XX
 DE Mycoplasma genitalium genome.
 KW M. genitalium; DNA; DNA gyrase; origin of replication;
 KM megabase shotgun sequencing method; open reading frame; ORF; ss.
 XX
 OS Mycoplasma genitalium.
 XX
 FH Key
 FT 8552..9184
 CDS
 FT /label= a
 FT /note= "Previously identified as MORF-20076, the
 FT encoded protein shows 27.59 percentage
 FT identity to thymidylate kinase (CDC8)
 FT from *Saccharomyces cerevisiae*"
 CDS
 FT 11252..12040
 FT /tag= b
 FT /label= MG009
 FT /note= "Previously identified as MORF-20078, the
 FT encoded protein shows 35.43 percentage
 FT identity to the *Bacillus subtilis* hypothetical
 FT protein covered in accession number
 FT GB:D26185_102"
 CDS
 FT 12069..12725
 FT /tag= c
 FT /label= MG010
 FT /note= "Previously identified as MORF-20079, the
 FT encoded protein shows 25.73 percentage
 FT identity to DNA primase (dnaE) from
 FT *Clostridium acetobutylicum*"
 CDS
 FT 14396..15217
 FT /tag= d
 FT /label= MG012
 FT /note= "Previously identified as MORF-20080, the
 FT encoded protein shows 31.50 percentage
 FT identity to the ribosomal protein S6
 FT modification protein (RimK) from *Escherichia*
 FT *coli*"
 CDS
 FT 16013..16013
 FT /tag= e
 FT /label= MG013
 FT /note= "Previously identified as MORF-19823, MORF-20080
 FT and MORF-20081, the encoded protein shows 33.04

FT percentage identity to 5,10-methylene-tetra-
FT hydrofolate dehydrogenase (foid) from E. coli"
FT
FT
FT CDS 17474..19243
FT /tag= f
FT /label= MG015
FT /note= "Previously identified as MORF-20084, the
FT encoded protein shows 32.23 percentage
FT identity to transport ATP-binding protein
FT (msbA) from E. coli"
FT
FT CDS 26478..27344
FT /tag= g
FT /label= MG023
FT /note= "Previously identified as MORF-20092, the
FT encoded protein shows 45.96 percentage
FT identity to fructose-bisphosphate aldolase
FT (tsf) from B. subtilis"
FT
FT CDS 27345..28448
FT /tag= h
FT /label= MG024
FT /note= "Previously identified as MORF-19826 and
FT MORF-20093, the encoded protein shows 46.84
FT percentage identity to GTP-binding protein
FT from E. coli"
FT
FT CDS 36987..38978
FT /tag= i
FT /label= MG032
FT /note= "Previously identified as MORF-20099, the
FT encoded protein shows 26.82 percentage
FT identity to ATP-dependent nuclease (adda)
FT from B. subtilis"
FT
FT CDS 39242..39904
FT /tag= j
FT /label= MG033
FT /note= "Previously identified as MORF-20100, the
FT encoded protein shows 35.90 percentage
FT identity to glycerol uptake facilitator
FT (gipe) from B. subtilis"
FT
FT CDS complement (39873..40514)
FT /tag= k
FT /label= MG034
FT /note= "Previously identified as MORF-20101, the
FT encoded protein shows 48.13 percentage
FT identity to thymidylate kinase (tdk)
FT from B. subtilis"
FT
FT CDS 40543..41787
FT /tag= l
FT /label= MG035
FT /note= "Previously identified as MORF-20102, the
FT encoded protein shows 30.71 percentage
FT identity to histidyl-tRNA synthetase (hiss)
FT from Mycobacterium leprae"
FT
FT CDS complement (44751..46277)
FT /tag= m
FT /label= MG038
FT /note= "Previously identified as MORF-20105, the
FT encoded protein shows 46.83 percentage
FT identity to glycerol kinase (gipk)
FT from E. coli"
FT
FT CDS complement (46268..47422)
FT /tag= n
FT /label= MG039
FT /note= "Previously identified as MORF-19831 and
FT MORF-20106, the encoded protein shows 43.20
FT percentage identity to glycerol-3-phosphate
FT dehydrogenase (GDR2) from S. cerevisiae"
FT
FT CDS 49377..49643
FT /tag= o
FT /label= MG041
FT /note= "The encoded protein shows 48.86 percentage
FT identity to phosphotransferase (ptsH) from Mycoplasma
FT capricolum"
FT
FT CDS 50060..51520

FT /tag= p
FT /label= MG042
FT /note= "Previously identified as MORF-19832 and
FT MORF-20108, the encoded protein shows 41.92
FT percentage identity to spermidine/
FT putrescine transport ATP-binding protein
FT (pta) from E. coli"
FT
FT CDS 51525..52382
FT /tag= q
FT /label= MG043
FT /note= "Previously identified as MORF-20110, the
FT encoded protein shows 26.51 percentage
FT identity to spermidine/putrescine transport
FT system permease protein (potB) from E. coli"
FT
FT CDS 52366..53220
FT /tag= r
FT /label= MG044
FT /note= "Previously identified as MORF-20111, the
FT encoded protein shows 29.45 percentage
FT identity to spermidine/putrescine transport
FT system permease protein C (potC) from E. coli"
FT
FT CDS 54658..55605
FT /tag= s
FT /label= MG046
FT /note= "Previously identified as MORF-20112, the
FT encoded protein shows 36.60 percentage
FT identity to sialoglycoprotease (gcp)
FT from Pasteurella haemolytica"
FT
FT CDS complement (56970..58310)
FT /tag= t
FT /label= MG048
FT /note= "Previously identified as MORF-19834,
FT MORF-20114 and MORF-20115, the encoded protein
FT shows 43.02 percentage identity to signal
FT recognition particle protein (fth) from B.
FT subtilis"
FT
FT CDS 58117..59079
FT /tag= u
FT /label= MG049
FT /note= "Previously identified as MORF-20114 and
FT MORF-20115, the encoded protein shows 44.78
FT percentage identity to purine-nucleoside
FT phosphorylase (deod) from E. coli"
FT
FT CDS 59083..59754
FT /tag= v
FT /label= MG050
FT /note= "Previously identified as MORF-20117, the
FT encoded protein shows 83.03 percentage
FT identity to deoxyribose-phosphate aldolase
FT (deoc) from Mycoplasma pneumoniae"
FT
FT CDS complement (64896..65731)
FT /tag= w
FT /label= MG056
FT /note= "Previously identified as MORF-20122, the
FT encoded protein shows 30.25 percent
FT identity to the protein disclosed in
FT GB:D26185_99 from B. subtilis"
FT
FT CDS complement (65713..66249)
FT /tag= x
FT /label= MG057
FT /note= "Previously identified as MORF-20123, the
FT encoded protein shows 38.90 percentage
FT identity to the protein disclosed in
FT GB:D26185_104 from B. subtilis"
FT
FT CDS 81047..82597
FT /tag= y
FT /label= MG067
FT /note= "Previously identified as MORF-19845, the
FT encoded protein shows 28.84 percentage
FT identity to glutamic acid specific protease
FT (SPase) from Staphylococcus aureus"
FT
FT CDS 91065..91919
FT /tag= z

```

FT      /label= MG070
FT      /note= "Previously identified as MORE-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"
FT

Query Match      3.9%; Score 124.6; DB 18; Length 580073;
Best Local Similarity 63.5%; Pred. No. 2.2e-16;
Matches 190; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY      2770 TTGGGCGCTTACCTCAGGCTAGTCCCTACAGCAATGTTGTTACATTAGTATCTT 2829
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      228703 TTTCGACACATCGTTCGACGATGCGTATCATAGTTCCTTACGTTGGATCTTG 228762
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      2830 GCATTTATGATCATCTTAGGATTAGCTATCGGATTCCTTAAGAGCTCAAGAAATTA 2889
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      228763 TTTATCTTTTACTCTTAGGACTCTGGATTTGGATCCCATGTACAGGGTAAAGAACTC 228822
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      2890 CAAGACAAAGGTTCAAAACAAACATTCAAAAAAGTTGATACCTTGACTGCTGTGGT 2949
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      228823 CAAGATGATCGTTTGTATGTCCTTAAAGGTTGATACACTCACAACTGCTGCGGT 228882
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      2950 TCAGTTTACAGAAAGTTATTACCCAACTGCTACAGTTAAGAAAAAAGCTGCTGTTT 3009
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      228883 ACTGTGTCAGAAAAAGATTATTACCCAACTGCTGTTGTAAGAAAAAGCACCTAGTGCATTG 228942
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      3010 GGTCGTGTAATCTGATAGAAAGAACTGCTGCTGCTAAGACCTGCTGCTCAGC 3068
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      228943 AAGCTGCTAATCTGATGTTAAAGAACTGCTGCTGCTTTTAAAGACCACTGTTCAACC 229001
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
AAQ81778
ID      AAQ81778 standard; DNA: 4182 BP.
XX
AC      AAQ81778;
XX
DT      09-AUG-1995 (first entry)
XX
DE      DNA encoding cytoadhesin protein.
XX
KW      Cytoadhesin protein; diagnostic probe; unusual codon usage;
KW      immunoassay; antigen; ds.
XX
OS      Mycoplasma gallisepticum.
XX
PH      key
FT      Location/Qualifiers
FT      CDS
FT      713..4081
FT      /tag= a
FT      /product= cytoadhesin protein
FT      misc_difference 887..889
FT      /tag= b
FT      /codon= seq:tga, aa:trp
FT      misc_difference 998..1000
FT      /tag= c
FT      /codon= seq:tga, aa:trp
FT      misc_difference 1484..1486
FT      /tag= d
FT      /codon= seq:cta, aa:lys
FT      misc_difference 1487..1489
FT      /tag= e
FT      /codon= seq:tta, aa:lys
FT      misc_difference 1559..1561
FT      /tag= f
FT      /codon= seq:tta, aa:lys
FT      misc_difference 1592..1594
FT      /tag= g
FT      /codon= seq:tta, aa:lys
FT      misc_difference 1676..1678
FT      /tag= h
FT      /codon= seq:tga, aa:trp
FT      misc_difference 1715..1717
FT      /tag= i

```

```

FT      /codon= seq:cta, aa:lys
FT      misc_difference 1853..1855
FT      /tag= j
FT      /codon= seq:tga, aa:trp
FT      misc_difference 1890..1892
FT      /tag= k
FT      /codon= seq:tga, aa:trp
FT      misc_difference 1904..1906
FT      /tag= l
FT      /codon= seq:tta, aa:lys
FT      misc_difference 1961..1963
FT      /tag= m
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2132..2134
FT      /tag= n
FT      /codon= seq:cta, aa:lys
FT      misc_difference 2237..2239
FT      /tag= o
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2273..2275
FT      /tag= p
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2417..2419
FT      /tag= q
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2504..2506
FT      /tag= r
FT      /codon= seq:tta, aa:lys
FT      misc_difference 2609..2611
FT      /tag= s
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2639..2641
FT      /tag= t
FT      /codon= seq:tga, aa:trp
FT      misc_difference 2858..2860
FT      /tag= u
FT      /codon= seq:tga, aa:trp
FT      misc_difference 3020..3022
FT      /tag= v
FT      /codon= seq:tga, aa:trp
FT      misc_difference 3083..3085
FT      /tag= w
FT      /codon= seq:tga, aa:trp
FT      misc_difference 3138..3140
FT      /tag= x
FT      /codon= seq:tga, aa:trp
XX
XX      US5378820-A.
XX      03-JAN-1995.
XX      PD
XX      PE
XX      09-NOV-1992; 92US-0973257.
XX      PR
XX      09-NOV-1992; 92US-0973257.
XX      PA
XX      (DOHM/) DOHMS J E.
XX      PA
XX      (KEEL/) KEELER C L.
XX      PI
XX      Dohms JE, Keeler CL;
XX      WP1: 1995-051314/07.
XX      DR
XX      P-PDSB; AAR64927.
XX
XX      Nucleic acid encoding cytoadhesin protein - used as a probe to
XX      diagnose Mycoplasma gallisepticum infection in poultry
XX
XX      Claim 1: Column 7-16; 13pp; English.
XX
XX      The sequence encodes a cytoadhesin protein from Mycoplasma
XX      gallisepticum. All or part of the sequence may be used as a probe
XX      for diagnosis of M. gallisepticum infection in poultry, e.g. turkey
XX      or fowl. The sequence has several UGA codons, which may be
XX      eliminated to avoid problems of expression in vitro due to irregular

```

Best Local Similarity	48.3%	Pred.No.	0.00044	
Matches	181;	Conservative	0;	Mismatches 194; Indels 0; Gaps 0;
2727	TAATTAAGACATATTAGTAACATCTCTATATACAACTTGGGGCCTTACCTTC	2766		
3033	TAGTCTCAAAATATATCTTATTAATTAATTTCTATTTTTCGTCGATTTAGTCAATGCGCTCA	3092		

2787 AUGGTAGCGCCACAGCAATTGGTAGTCATTAAGTATTTTGCAATATGATCATCTT 2846
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3093 TGAATCTTCTCCTACTATTTCCTATTCCTCCATGTATTAATGCATATAAATGGTTTT 3152

2847 AGAGTTACGTAFCGGATTCCTTAAAGACTCAAGAATAATTACAAGCANAAGGCTTCA 2906
||||| ||||| ||||| | ||||| ||||| |||||
3153 AGCTTGCGAGTATGTGGGATTCCTAATGGCTTAACATATAAANAGCATATTAAGCTGGTTTTGA 3211

2907 AACACATTCAAAAAGTTGATACCTTGGACTGCTGCTGTTGAGTTACATTACAAGAAGAT 2966
 2927
 3213 ATTGCACACGCTAAAGTTGGACATTCGACTCAGCTGTGGTGGTGCTTTTCAAAAAGAT 3277

2967 TATTACCAACTGCTAACGTTTAAAGAAAAACCTGCTGTAGTGCTGAATTCGG 3028
||||| | ||| ||| ||| |
3273 TATTGCATTAACAATTCATAATATGTGAATCTTAACCAACAATGTTAAGCGTCGAGC 333

3027 TGATAGAACCTGCTGCTGCTGCTAAACCTGCTGCTCAGGTAAACCATCTGCACCAA 308
| | | | | | | | | |
3333 TAAATTAACCAATAGTTCACCTGCTAATCTCAATTACAAMTGATCTGGTTTCTAG 339

3087 AGCTAGCTCACCAGC 3101
 | | | | | | |
 3393 ACCGACCTCCACCATTC 3407

RESULT 6
0075107/c

ADQ75017c
ID ABQ75107 standard; cDNA; 4985 BP.
XX
AC ABQ75107;
XY

AA	
DT	01-NOV-2002 (first entry)
XX	
DE	Anopheles gambiae odourant receptor 2 genomic DNA seq ID NO:10.
VV	

AA Anopheles gambiae; mosquito; olfactory gene; arrestin 1; pest control,
KM odourant receptor; olfaction; gene; ds.
XX
Anopheles gambiae.

US	anoplastic gamma-ray
XX	
XX	
PN	WO200259274-A2.
XX	
01-AUG-2002	
en	

XX 28-TAN-2002; 2002MO-US02549.
 PF
 XX 36-TAN-2001; 2001JUS-264649P.
 XX

PR 20-JAN-2001; 200205-0056405.
PR 24-JAN-2002; 200205-0056405.
XX
XX
PA (UYVA-) UNIV VANDERBILT.

AA
PI Zweibel LJ;
XX
DR WPI: 2002-627421/67.
 WPI: 2002-627421/67.
 WPI: 2002-627421/67.

DR P-PSDB; AB024000.

XX New mosquito olfaction polypeptides and polynucleotides, useful for

PT mosquito management, i.e. controlling the pest and disease vectors, or

PT mosquito management, i.e. controlling the pest and disease vectors, or

XX *Disclosure; Fig 4a; 96pp: English.*

CC The present invention uses a P-loop amino acid sequence comprising a 383, 394, 380 411, 412, 391, 157 or 401 resid
CC polypeptide comprising a 383, 394, 380 411, 412, 391, 157 or 401 resid
CC amino acid sequence (see Abp52833 to Abp52840) (S1), a conservatively
CC modified amino acid sequence of them, or a sequence of (S1) with at lea

CC 20 consecutive residues. Also described: (1) an isolated polynucleotide comprising: (a) a nucleotide sequence encoding the purified Anopheles gambiae olfaction polypeptide; or (b) a nucleotide sequence that hybridises under stringent conditions to a hybridisation probe comprising a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (52), or its complement; and (2) a method for identifying an agent that binds to mosquito olfaction molecules comprising: (a) providing an isolated mosquito olfaction molecule; (b) contacting a test agent with the isolated mosquito olfaction molecule; and (c) detecting specific binding of the test agent to the isolated mosquito olfaction molecule, where the presence of specific binding identifies the test agent as a mosquito olfaction-binding compound. The mosquito olfaction molecules are useful for mosquito management, i.e. controlling this pest and disease vector. A method from the present invention of screening for substances that modulate arrestin-odourant receptor interaction is useful for identifying pest control agents. The present sequence represents Anopheles gambiae odourant receptor 2 genomic DNA from the present invention.

CC N.B. The features given in figure 4a are tentative and do not directly encode SEQ ID NO:6 (ABP52835).

CC Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 other;

XX Query Match

Best Local Similarity 2.0%; Score 62.6; DB 24; Length 4985;

Matches 333; Conservative 0; Mismatches 396; Indels 3; Gaps 3;

OY 253 ATCAATTTAGATGTTTACTTAACCATTTATGTTTGAATCTAAGTATGTTGGT 312

DB 3070 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3011

OY 313 GGATCAAAAGTAAACAAATAGTTTCA-GATTACACACTGACAGAAATAGATTGATCA 371

DB 3010 CACACACACAAACAAATTAATAGAAATTAATTAATTAATTAATTAATTAATTAATTA 2951

OY 372 AAGACAAACAGACATATTATGCTGTGTTGTTAATGATGAGCTAACGTTTCAAT-TTAA 430

DB 2950 TAACATATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2891

OY 431 AAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 490

DB 2890 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2831

OY 491 TTGTGTGTGTATATATCCAGCTCAAGTATTAATTAATTAATTAATTAATTAATTAATTA 550

DB 2830 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2771

OY 551 ATTTTCAAAACCAAACTCAAGGTGAATTTGTTAATGACTTCAATTTAGTGGCCAACT 610

DB 2770 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2711

OY 611 TTAC-CTAAGATTTTACACCAGATGCTTAACCTTACATTAATTAATTAATTAATTAATTA 669

DB 2710 AACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2651

OY 670 AATGACGTCAACACTGCACTGTGCTTCCCTGGCCAGTGAAGTGAAGTGAAGTGAAGTGA 729

DB 2650 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2591

OY 730 GATGATGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 789

DB 2590 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2531

OY 790 ACTTAACCACTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 849

DB 2530 AGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2471

OY 850 GCAAAACATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 909

DB 2470 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2411

OY 910 TTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969

DB 2410 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2351

OY 970 ACATTAGCTACT 981

DB 2350 AATTTCTGTAAT 2339

RESULT 7

ABL33825/C

ID ABL33825 standard; DNA; 8771 BP.

AC ABL33825;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1798.

XX Human: immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antihaemic; cytosolic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001: 2001MO-EP07537.

XX 30-JUN-2000: 2000DE-1032529.

XX 01-SEP-2000: 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI: 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -

XX Claim 1: SEQ ID NO 1798; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 8771 BP; 2734 A; 59 C; 1660 G; 4318 T; 0 other;

XX Query Match 1.9%; Score 60.8; DB 24; Length 8771;

XX Best Local Similarity 46.3%; Pred. No. 0.0039;

XX Matches 200; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

OY 256 AATTTAGATGTTTACTTAACCATTTATGTTTGAATCTAAGTATGTTGGTGGCA 315

DB 5053 ATTAATCTATACATACATATAAATAATTAATTAATTAATTAATTAATTAATTAATTA 4994

OY 316 TACAAGTAAACCAATAGTTTCAGATTACACACTGACGAAATGATGATGATGATGATGATGAT 375

DB 4993 TACTACACAAAAAACCCTTAATAAACATTTATGCAAAATTAATAAACCAATTAACAAAA 4934

OY 376 CAAACAAGACATATTATGCTGCTGTTGTTAATGATGAGCTTAACGTTCAATTTAAAAAGA 435

	RESULT 8
AAA030290/c	
ID AAA030290 standard; DNA;	3489 BP.
XX AC	
XX AA030290;	
XX DT	11-SEP-2000 (first entry)
DE XX	Kaposi's sarcoma-associated herpesvirus LANA gene.
KW KW	Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus; latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KW KM	Human herpes virus 8; HHV8; rhadno virus cis-acting element; RVCAG;
KX OS	Kaposi's sarcoma; primary effusion lymphoma; PEL; human immunodeficiency virus; HIV; multicentric Castleman's disease; ds
XX KS	Kaposi's sarcoma-associated herpesvirus.
FH PH	Key Location/Qualifiers
CDS FT	1..3489
FT FT	/tag= a
FT FT	/product= "LANA"
FT misc_signal	40.. .50
FT /tag= b	
FT /note= "nuclear localisation signal, NLS"	
FT misc_signal	190.. .210
FT /tag= C	
FT /note= "nuclear localisation signal, NLS"	
XX MO	MO200029626-AI.
XX PD	25-MAY-2000.
XX PE	19-NOV-1999; 99WC-US27508.
XX PR	19-NOV-1998; 98US-0109422.
PR PA	21-APR-1999; 99US-0298568.
XX PA	(KIEF/) KIEFE E D.
PA (BALI/)	BALLESTAS M E.
PA (KAJE/)	KAYE K M.
PI Kieif ED,	Ballestas ME, Kaye KM;
XX WPI:	2000-387829/33.
DR P-PSTD;	AAY96255.
PT Treating or preventing a disease associated with rhodino virus infection in a mammal which includes Kaposi's Sarcoma and Primary	

PT Effusion Lymphoma -
XX
PS Disclosure; Fig 6; 70pp: English.
XX
XX The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC (KSHV) latency-associated nuclear antigen (LANA) gene. KSHV is also known
CC as Human Herpes Virus 8 (HHV8), and belongs to the rhadno virus, or
CC gamma-2 herpes virus class. The LANA protein is necessary for the
CC efficient persistence of rhadno virus DNA in mammalian cells. Persistent
CC rhadno virus infection is implicated in a variety of diseases e.g.,
CC Kaposi's Sarcoma (KS). Primary Effusion Lymphoma (PEL) and multicentric
CC Castleman's disease. In addition, KS is a common malignancy in HIV
CC patients. KSHV persists in host cells in a latent form. One of the few
CC genes expressed from the latent viral DNA is LANA. LANA associates with
CC both human chromosomes and with the rhadno virus cis-acting element
CC (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC "tied" to the host chromosomes. This allows the viral DNA to persist in
CC the host cell. The present sequence may be used to screen and identify
CC molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC with the latency cycle of this virus. Potential antiviral treatments for
CC the above mentioned diseases may therefore be based on LANA deregulation.

SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;

[illegible]

RESULT 9	
ID	AAF82901/C
XX	AAF82901 standard; DNA; 3489 BP.
XX	
XX	AAF82901;
XX	
DT	29-JUN-2001 . (first entry)
XX	
DE	Nucleotide sequence of KSHV tethering protein, LANA.
XX	
XX	Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW	Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8.
KW	KSHV; latency-associated nuclear antigen; LANA; ds.
XX	
OS	Kaposi's sarcoma associated herpesvirus.
XX	
FH	Key
FH	1..3489
FT	Location/Qualifiers
FT	/*tag= a
XX	
PN	W0200125484-A2.
XX	
XX	12-APR-2001.
PD	
XX	
PF	29-SEP-2000; 2000WO-US26908.
XX	
PR	01-OCT-1999; 99US-0410399.
XX	

PA (UNMI) UNIV MICHIGAN.
XX
XX
PI Robertson ES, Colter MA;
XX
XX WPI: 2001-281736/29.
DR P-PSDB: AAB62331.
XX
XX
PT A composition for use in gene therapy comprises an expression vector
PT that includes a nucleic acid sequence encoding a nucleic acid binding
PT protein -
XX
XX
PS Disclosure: Fig 9A; 60pp; English.
XX
XX
CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the nucleotide sequence of the Kaposi's sarcoma
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC antigen (LANA), which acts as a tethering protein.
XX
XX
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
Query Match 1.9%; Score 60.2; DB 22; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.004;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;
OY 2998 CCTGCTGTTTAAAGTGGTGAATCTGATTAAGAAACCTGCTGCTCTAAACCT 3057
DB 2236 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2177
OY 3058 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3115
DB 2176 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
OY 3116 -CTAATCTGGTGGCGCTTACAAACCAACCTGCTGCTGCTGCTGCTGCTGCT 3174
DB 2116 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2057
OY 3175 GCTCC 3179
DB 2056 GCTGC 2052
RESULT 10
ABA93487/C
ID ABA93487 standard; DNA: 3489 BP.
XX
XX ABA93487;
AC
XX
XX 25-APR-2002 (first entry)
DT
XX
XX Kaposi's sarcoma-associated herpesvirus LANA protein encoding DNA.
DE
XX
XX Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
KW KSHV terminal repeat; rhadino virus cis acting element; episome;
KW primary effusion lymphoma; latency-associated nuclear antigen;
KW gene therapy; gene transfer; gene; ds.
XX
XX
OS Human herpesvirus 8.
XX
XX
FH Key 1.3489 Location/Qualifiers
FH CDS
FT /tag= a
FT /product= "LANA protein"
FT /note= "latency-associated nuclear antigen"

XX
XX US6322792-B1.
PN
XX
XX 27-NOV-2001.
PD
XX
XX 21-APR-1999; 9905-0298568.
PF
XX
XX 19-NOV-1998; 9805-109422P.
PR
XX
XX (KIEF/) KIEFF E D.
PA
XX
XX Kieff ED, Balleskas ME, Kaye KM;
PI
XX
XX WPI: 2002-153769/20.
DR P-PSDB: ABB05621.
XX
XX
PT System for episomal retention of plasmids in mammalian cells, useful in
PT gene therapy, comprises rhadinoviral LANA and RVCAE sequences -
XX
XX
PS Claim 1; Fig 6; 27pp; English.
XX
XX
CC The present invention describes a system (A) for maintaining a plasmid
CC as an episome in mammalian cells, comprising the rhadinoviral sequence
CC LANA (latency-associated nuclear antigen) of 3489 base pairs (see
CC ABA93487, S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC present in the plasmid. Also describes is a method for maintaining a
CC closed circular DNA in a cell by expressing (S1) in the cells and having
CC (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC particularly used in gene therapy (or other gene transfer applications)
CC that uses mammalian cells in which LANA is expressed. (A) improves
CC persistence of gene therapy vectors in cells. The present sequence
CC encodes Kaposi's sarcoma-associated herpesvirus (KSHV, also called human
CC herpesvirus 8) LANA protein, which is used in the exemplification of the
CC present invention.
XX
XX
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
Query Match 1.9%; Score 60.2; DB 24; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.004;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;
OY 2998 CCTGCTGTTTAAAGTGGTGAATCTGATTAAGAAACCTGCTGCTCTAAACCT 3057
DB 2236 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2177
OY 3058 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3115
DB 2176 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
OY 3116 -CTAATCTGGTGGCGCTTACAAACCAACCTGCTGCTGCTGCTGCTGCTGCT 3174
DB 2116 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2057
OY 3175 GCTCC 3179
DB 2056 GCTGC 2052
RESULT 11
AAV73805
ID AAV73805 standard; DNA: 32207 BP.
XX
XX
XX AAV73805;
AC
XX
XX 25-FEB-1999 (first entry)
DT
XX
XX KSHV LUR DNA (nucleotides 105,301-137,507).
DE
XX
XX Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW dihydropyrimidine reductase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;

Key	Location/Qualifiers
XX OS	Kaposi's sarcoma-associated herpes virus.
XX KH	
XX FH	
XX CDS	1142..2794
XX FT	/*tag= a
XX FT	/product= complement-binding protein
XX FT	8699..11236
XX FT	/*tag= b
XX FT	/product= glycoprotein B
XX FT	complement (17261..17875)
XX FT	/*tag= c
XX FT	/product= interleukin 6
XX FT	complement (21548..21832)
XX FT	/*tag= d
XX FT	/product= macrophage inflammatory protein II
XX FT	complement (27137..27424)
XX FT	/*tag= e
XX FT	/product= interferon regulatory factor 1
XX FT	2861..29741
XX FT	/*tag= f
XX FT	/product= protein T1.1
XX FT	complement (58976..60175)
XX FT	/*tag= g
XX FT	/product= glycoprotein M
XX FT	complement (69412..69915)
XX FT	/*tag= h
XX FT	/product= glycoprotein L
XX FT	complement (88410..88910)
XX FT	/*tag= i
XX FT	/product= interferon regulatory factor 2
XX FT	89600..90541
XX FT	/*tag= j
XX FT	/product= interferon regulatory factor 3
XX FT	90173..90643
XX FT	/*tag= k
XX FT	/product= glycoprotein X
XX FT	complement (93636..94127)
XX FT	/*tag= l
XX FT	/product= interferon regulatory factor 4
XX FT	complement (111931..112443)
XX FT	/*tag= m
XX FT	/product= capsid protein IV
XX FT	complement (123808..127296)
XX FT	/*tag= n
XX FT	/product= immediate early protein
XX PN	WC9804576-A1.
XX PD	05-FEB-1998.
XX PE	22-JUL-1997;
XX PE	97WO-US13346.
XX PR	29-NOV-1996;
XX PR	96US-0757669.
XX PR	25-JUL-1996;
XX PR	96US-0688243.
XX PR	25-JUL-1996;
XX PR	96US-0688349.
XX PR	25-JUL-1996;
XX PR	96US-0686350.
XX PR	25-JUL-1996;
XX PR	96US-0687253.
XX PR	25-JUL-1996;
XX PR	96US-0688814.
XX PR	05-SEP-1996;
XX PR	96US-0708678.
XX PR	10-OCT-1996;
XX PR	96US-0728323.
XX PR	13-NOV-1996;
XX PR	96US-0747887.
XX PR	13-NOV-1996;
XX PR	96US-0746640.
XX PA	(UYCO) UNIV COLUMBIA NEW YORK.
XX PI	Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;

DR WPI: 1998-130615/12.

XX New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients

XX Example 2: Page 135-203: 230pp; English.

CC This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (e) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection.
CC While the protein can be used in protective vaccines, Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, leukaemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. diphthololate reductase gene
CC can be inhibited with mifepristone. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP II may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.

XX Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

XX Query Match 1.9%; Score 60.2; DB 19; Length 137507;

XX Best Local Similarity 61.6%; Pred. No. 0.012;

XX Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

XX 2998 CCTGCTGCTTTAGTGGTGGTAAATCTGATTAAGAAACGCTGCTGCTAAACCT 3057

DB 125061 CATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125120

QY 3058 GCTGCTCAGCTAAACATCTGACCAAAAGCTAGCTACCAAGCTAAACACTGGGC-- 3115

DB 125121 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125180

QY 3116 -CTAATCTGGTGGCGCTACAAACCAACTGCTCTTAACCGAGCTGCTCAAAACCAACC 3174

DB 125181 GCTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125240

QY 3175 GCTCC 3179

DB 125241 GCTGC 125245

RESULT 13

AB042292/c

ID AB042292 standard; DNA; 535 BP.

XX AB042292;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 28883.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

PN WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guelig D;

XX WPI: 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

XX Claim 12; 56pp + sequence listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 535 BP; 98 A; 29 C; 55 G; 353 T; 0 other;

XX Query Match 1.9%; Score 59; DB 24; Length 535;

XX Best Local Similarity 48.6%; Pred. No. 0.0041;

XX Matches 161; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

XX 1203 GAATCAAGTTAAACAGATGATATTTGAAGCTTCAAGCACTGATTAACGATCAACAA 1262

DB 467 GAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 408

QY 1263 CGAACAACGACACAGCTGATCATCTAGTGTTCACAGAGTGTGGAAACAGAAATAC 1322

DB 407 TAAACGAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAAT 348

QY 1323 TACTACACTTTCGCAACAGTTTCTTAATCTTAAATCTTAATCTTAATCTTAATCTTAAT 1382

DB 347 TAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 288

QY 1383 TGATAGTAACCAACTTCTGCAACCAAAATAGATGAAGTAAATTTGGCGAGATCTAACGT 1442

DB 287 TACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 228

QY 1443 TATTGAACGACAGATATATGCTGGAATACAGATTTAGTATTTCAAAATGAATTCATAC 1502

DB 227 AACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168

QY 1503 TATGACGAGAACTTATTCGAAACACAAAT 1533

DB 167 TATTAACGACGAAATTAATAATAATAAT 137

RESULT 14
 ID ABO42293 standard; DNA: 535 BP.
 AC ABO42293;
 DT 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28884.
 XX Human: cytosine methylation: 5'-CpG-3'; uncil: cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 OS Homo sapiens.
 PN WO200218632-A2.
 PD 07-MAR-2002.
 PE 01-SEP-2001; 2001WO-EP10074.
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 PA (EPIC-) EPICGENOMICS AG.
 PI Olek A. Piepenbrock C. Berlin K. Guettig D;
 DR WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 XX Claim 12; 56pp + Sequence Listing; 56pp: German.
 PS This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to both classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridization to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridized to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABO1310-AB054121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX Sequence 535 BP; 353 A; 55 C; 29 G; 98 T; 0 other;
 SQ
 Query Match 1.94; Score 59; DB 24; Length 535;
 Best Local Similarity 48.6%; Pred. No. 0.0041;
 Matches 161; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
 OY 1203 GAATCAAGTTAAACAGATGATATGAGCTTAAACAGTGAACGGTACAAACAA 1262
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 69 GAATTAATAATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 128
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1263 CGGACACGACACAGCTGATCATCTAGTGGTCAACAGTGTGGAACAGGAATTC 1322
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 129 TAAACGATACGATACGATACGATACGATACGATACGATACGATACGATACGAT 188
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 1323 TACTAACCTCTGCAACACTTTCTTACTTAACTTAACTTAACTTAACTTAACTTAACT 1382
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 189 TAAATATCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 248
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1383 TGATAGTAAACCAACTCTGCAACAAATATGATGAACACTTATGGGCGAGTCTTAAGT 1442
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 249 TACGATTAATAATCTATATATAACGATTAATAATAATAATAATAATAATAATAATA 308
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1443 TATTGACGACGAATATATGCTGATACATATAGTATGCAATATGCAATATAC 1502
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 309 AACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 368
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1503 TAATGACGAAACTTATCCGAACACAACTT 1533
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 369 TAATTAACGACGAACGAATTAATAATAATAATAATAATAATAATAATAATAATA 399
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
 ID ABA92787 standard; DNA: 640681 BP.
 AC ABA92787;
 DT 27-MAR-2002 (first entry)
 DE Buchnera sp. genomic DNA SEQ ID NO:1.
 XX Buchnera: cockroach-symbiotic bacterium; cockroach extermination;
 XX circular; ds.
 OS Buchnera sp.
 PN JP2001292771-A.
 PD 23-OCT-2001.
 PE 07-APR-2000; 2000JP-0107160.
 PR 07-APR-2000; 2000JP-0107160.
 PA (RIKA) RIKAGAKU KENKYUSHO.
 DR WPI; 2002-126043/17.
 XX A genomic DNA of cockroach-symbiotic bacterium
 PS Claim 1; Page 16-230; 237pp; Japanese.
 CC The present invention describes a gene (I) derived from Buchnera sp.
 CC containing the DNA (a) or (b); (a) has a fully defined base pair
 CC sequence selected from a table of sequences found in the Buchnera sp.
 CC genomic DNA of ABA92787 given in the specification or is a DNA selected
 CC from complementary DNA sequences, and (b) is a DNA which hybridizes with
 CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
 CC vector (II) containing (I); (2) a transformant (III) containing (II);
 CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
 CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
 CC (d); (c) is a DNA containing a fully defined sequence with a DNA; and (5) a
 CC or ABA92789 and (d) is a plasmid which hybridizes with a DNA; and (5) a
 CC method for the preparation of a fully defined sequence with a DNA; and (5) a
 CC the expression protein of the objective protein is collected from the
 CC resultant culture. The DNA is useful for developing agricultural
 CC chemicals for exterminating cockroaches. The present sequence represents
 CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
 CC present invention.
 XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
 SQ
 Query Match 1.88; Score 58.4; DB 24; Length 640681;
 Best Local Similarity 47.3%; Pred. No. 0.047;
 Matches 209; Conservative 0; Mismatches 231; Indels 2; Gaps 1;
 OY 241 AATGACACAAATTTCAATATAGATAGTTTAAACCACTTATATGCTTATAGATCTAAGT 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
Db 525047 ATTAAAAAATAAACACAGAAATTTAAATACATACCATTTCTAAAAAGTAATTTATT 525106
QY 301 GATGATTGGTGATACAAAGTAACAAATAGTTTCAGATTACACACTAGCAGAAAT 360
Db 525107 TTAATAAAGAAAGAAAAATATTAATAAATTAATTCACAAAAAGAAATTAATAAAAAA 525166
QY 361 AGATTGATCAAAAGACAAACAGAGCATATTATGCTGTGGTAAATGATGAAGCTAAC 420
Db 525167 GCAATTGTAAATAGATCCTTCATCTATAAAAAAATTAACAATCAAGAAAGCACA 525226
QY 421 GTTCATTAAAAAGAAATTAATACTCAATCAATAGAAATGTAATAGAAACAAATTC 480
Db 525227 ATTTATTTCAAAAAAATCAAGATTAATTTTATATTCAGAAAAATTTAAATTAATTT 525286
QY 481 AAGTTGTAATTGGTGTGTGATATCCAGCTCAGCTAATTAGATTACTGATGATGG 540
Db 525287 GTTGAATTAATAAAGCTGATTAACCTTAAATTAACATTGCGAAAAATTAAGAGATTTATGATGG 525346
QY 541 ACTAATTTAATTTTACAAACCAAGCTCAAGGTAAGTGTAAATGACTTCATTTAGAT 600
Db 525347 TATATAGAAATATTAC--ACAATACTCAACTAAAGAAAAAGAGATTAATAGTATTTC 525404
QY 601 GCGCCATCTTACCTAAAGATTTCACCCAGATTGGTATTAATTAATTAATTAATTAAT 660
Db 525405 AAGTCAAAATATAACAACAAGCATATCGATTATCTAGATTGCAATTAATACACCAAG 525464
QY 661 ATCTTACCAATGACGTCACA 682
Db 525465 ATTTTCAAAAATAGCTCAGA 525486
```

Search completed: October 10, 2003, 08:36:31
Job time : 805 secs

THIS PAGE BLANK (USTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 07:41:45 : Search time 6303 seconds
(without alignments)
12296.839 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189

Sequence: 1 atgaatattctcaaaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estma:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_iny:*
19: em_gss_pln:*
20: em_gss_vrl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_pro:*
24: em_gss_rod:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	2.2	1101	29	CNS0039G
2	67	2.1	1200	13	BX437758
3	66.4	2.1	1101	29	CNS00EVL
4	65.8	2.1	844	29	BX139987

Result No.	Score	Query Match	Length	ID	Description
5	65.2	2.0	1101	29	CNS0039G
6	64.6	2.0	829	29	BX173672
7	64.2	2.0	773	29	CNS01WTC
8	63.8	2.0	1200	13	BX447454
9	61.8	1.9	470	29	FR0018463
10	61.8	1.9	480	29	BZ643398
11	61.8	1.9	843	29	BZ643413
12	61.8	1.9	1225	29	CNS0161D
13	61.6	1.9	735	29	CNS04NSM
14	60.6	1.9	641	28	AQ946120
15	60.2	1.9	1101	29	CNS00EVE
16	60	1.9	941	13	BUI26589
17	60	1.9	1056	13	BX415058
18	59.8	1.9	450	29	FR0025683
19	59.8	1.9	488	12	BJ396861
20	59.8	1.9	575	29	BZ780846
21	59.8	1.9	615	28	BH766948
22	59.8	1.9	994	13	BX414650
23	59.2	1.9	619	29	FR0006944
24	59.2	1.9	1201	13	BX458623
25	59	1.9	317	12	BJ328366
26	58.8	1.8	619	29	FR0013713
27	58.8	1.8	759	29	CNS06QXV
28	58.8	1.8	813	29	BX143161
29	58.8	1.8	836	28	BH491109
30	58.8	1.8	1007	29	CNS06X9S
31	58.4	1.8	1126	13	BX446391
32	58.2	1.8	332	14	C92788
33	58.2	1.8	641	12	BM181884
34	58	1.8	257	29	FR0026883
35	57.8	1.8	1101	29	CNS0100X
36	57.4	1.8	1101	29	CNS00ESI
37	57.2	1.8	427	29	BZ422321
38	57.2	1.8	649	12	BM160056
39	57.2	1.8	1101	29	CNS001FB
40	56.8	1.8	639	29	CNS038CX
41	56.8	1.8	989	29	CNS02H44
42	56.6	1.8	1101	29	CNS00EVL
43	56.4	1.8	596	29	FR0019514
44	56.4	1.8	1201	13	BX461105
45	56.2	1.8	501	29	FR0048173

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS
DEFINITION
BACR08K10 of RCT-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
ACCESSION
AL063921.1 GI:4941778
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; *Drosophilidae*; *Drosophila*.
1 (bases 1 to 1101)
Genoscope.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley *Drosophila* Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of

FEATURES

```
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
7007
```

BASE COUNT	ORIGIN	201 a	64 c	131 g	202 t	503 others

Query Match	2.28;	Score 70.8;	DB 29;	Length 1101;	
Best Local Similarity	16.3%;	Pred. No. 0.0024;			
Matches	94;	Conservative	225;	Mismatches	227;
				Indels	0;
				Gaps	0;

[illegible]

```

QY      67 GGTTCGCAAGCTTTGGCITTAAGCATAGATAAGATACGATTACACAGCAATAGAT
          :: |:: ||:: ::: ::: |:: |:: ::: ::: .....
Db     991 DRDAAATADAGGRDRGRKKKDKDDDDKCGKRKKAKAANKTKMWDMDMDKD

```

OY 127 AATCAAGCAGAACCGCTAGTGTCTATTGTTTAAGACTGGAGGCTCTTCGAAATTAACGC
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 931 KWMGSAKDRKADDDDGAGDKDDGCKGAKDADDPTDGTGKKDDDKDWDMWRAKKGWGATN 872

266

OY 187 TCGTTGTCATACAGTCTTAGAGATGGTGATGAATCACTTAATAACGCCGCATAC
|::: : | ::::: |::: :::::: |: ::||: : : :
Db 871 AAAAAIDMWWGCHADADWTINDAADDWMAADD RNDAAWAWMWDMAGATRDARDKWDRA 812

[illegible]

OY 307 TGTGTCGATACAAAGTAAACAATAGTTTCAGATTACCAACACAGCAGCGAATGGATTA
:: :: :: :: | :: :: :: :: | :: :: :: :: |
Db 751 WDQDDDDDRDAGCTAGKRWRTWKRRMKRDTFMDADDTTRDDRRRRRGDGADAKG 69

DB

691 KTTGRRRRRDRATMDRIDAMADAAWTTTDTDDMCKRDRRRRGARRRRRTTARAAM 63

QY 427 TTAAGAATTATCTACCAACATGTGGTAAAGGAAACAACATTTTCAGG
::: | : :::: | : : | : : : : : : : : : : : : : : : :
Db 631 DWMTWAKWDKWTADRDNDRAAATLWIDARKDRDAKAKAPARDRRARAADR

OY 487 GTAATTCGCTGTATTATAACCACTCAGCAATTAAGATTACCGAAGCATGGAAATGAAGA
:: : | | | | - | | | : ||| : : : |
Db 571 RWTTRKGTATTTTMAAPAAWMAWMAWMTTATTTTWTTTTTTTTTTTTTTTTTTTTAA 51

OY 547 TTTAATTTTACAAACCAGCTCAGTGTAATTGTT 582
::|::| |::|::|::|::|::|
Db 511 WNNAAWWTATWAAWTTAAANAAAAAAMATTTTTTT 476

	RESULT 2	
	BXA37758	
	pv137758	
	1200 bp	
	mRNA	linear
	EST_15-MAY-20	
	cccccccccgcac	

LOCUS	BX437758
DEFINITION	Homo sapiens THYMUS Homo sapiens cDNA clone U56A09v01B
ACCESSION	BX437758 1 GI:30773605
	U56A09v01B BX437758
	5-PRIME, mRNA sequence.

VERSION	BA43/10.1
KEYWORDS	EST.
SOURCE	Homo sapiens. (human)

BP 191 91006 Evry Cedex
Email: segretef@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen Contact : Feng Liang Email : fliang@lifetech.com URL :
Invitrogen 1600

FEATURES
Invitrogen
http://fulllength.invitrogen.com/InvitrogenCorporation
Faraday Avenue Genoscope sequence ID : CS0CAP008CA010P1
location/qualifiers
1..1200

...nos

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon..9606"
/clone="CS0CAP008YB01"
/tissue_type="THYMUS"
/clone_id="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT-6; 1st strand cDNA was primed
with a NotI-oligo (dt) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned in
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
71 b 310 t 274 others

```

BASE COUNT	515 a	30 c	12 9	22 5
ORIGIN				
Query Match	2.18;	Score 67;	DB 13;	Length 1200;
	30 aa;	Pred No. 0.014;		

	Best Local Similarity	30.7%	Mismatches	440;	Indels	13;	Gaps	
	Matches	283;	Conservative	181;				
QY	13	AAAAACTTAAAGTTATACATTGATTAGCTGCATTAGCTGTATTGGAGCCTTGTCCT	7					

D6 AAAAAAAAAAAAAAAWMAATWWMAAARRTDGAATKADDAITGGGGGMMTTTTTMMWTAA 3
73 GCAGCGTTGGCTTTAGCAATCAGATAAGACTAACGATACACGCCAATTAGTAATCAA 1

```

Db      : :: ||| |||||
326  RRDKTTTTTTTTTMAAARATAKRGNGADANTAMWTTWTTTATMMWMDDBDAAK 3
133  GCAAGAAGCTAGATGCTAATTCGTTAGACTTCGAGGCTTCGACAAATAGTTCGTG 1
07      : :: ||| ||||| : :: ||| ||||| : :: ||| ||||| : :: ||| |||||

```

Db 386 AMADAGATATTTT...
193 TTCATACAGTCCTAGACATGTCGATTAACCTTATACAGCAGCTAATGAGCAAT 2

[illegible]

```

Db      501  MWAARGRARGATTTTCTTTTAAWWMGAGAGTAAWTTTCTTTTCTTTTAAWMAATWT
QY      313  GGATCAAGAGTAAACAATAG-TTTGAGATTACACAACTAGCAGAAATAGATTGGATCA

```

[illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32			
Db	621	A	A	R	G	M	K	T	W	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
QY	432	A	G	A	T	T	A	T	A	C	T	A	C	T	A	A	T	T	G	T	A	T	T	A	G	A	A	C	A	C	A	T	T	C	T

Db 681 AAMAMGAARRRATTTWTAAMAMWMAAANAAGACAGANAATTTTTTTATKACARG
 492 TGGTGGTGTGTAATACAGCTCACGTAATTAGATTTTCATGATGGACTAATTTAA

Db 741 ATTTTAMWMTATRAAGARGADTTTTTATATMTATTTTTTMMANAAGATGKAAAAAAMW
 552 TTTTACAACCAACTCAAGTCMAATTCATATACCTCAATTTAGATGGCCAAATCTT
 OY

Table 1

Variable	Mean	SD	Min	Max
Age	67.8	9.0	50	85
Gender	Male			
Marital status	Married			
Educational level	High school			
Occupation	Retired			
Income	\$1,200			
Health status	Good			
Social support	Low			
Stress level	High			
Mental health	Fair			
Physical activity	Low			
Dietary habits	Poor			
Tobacco use	No			
Alcohol consumption	Occasional			
Chronic diseases	Hypertension			
Medication adherence	Low			
Quality of life	Low			
Life satisfaction	Low			
Depression symptoms	Present			
Anxiety symptoms	Present			
Cognitive function	Normal			
Memory recall	Good			
Attention span	Short			
Executive function	Impaired			
Problem solving	Difficult			
Decision making	Slow			
Emotional stability	Unstable			
Resilience	Low			
Coping strategies	Avoidance			
Help-seeking behavior	Low			
Social engagement	Low			
Volunteering	No			
Religious participation	No			
Family size	Small			
Childhood trauma	No			
Adult trauma	No			
Current stressors	Financial			
Health concerns	Chronic pain			
Relationship issues	No			
Work-related stress	No			
Environmental factors	Noise pollution			
Access to healthcare	Good			
Insurance coverage	Private			
Transportation access	Good			
Community resources	Limited			
Neighborhood safety	Safe			
Crime rates	Low			
Police presence	High			
Public services	Good			
Local government	Responsive			
Political participation	No			
Voting behavior	No			
Civic engagement	Low			
Trust in institutions	Low			
Perceived corruption	High			
Economic outlook	Pessimistic			
Future prospects	Uncertain			
Hope for the future	Low			
Optimism	Low			
Positivity	Low			
Gratitude	Low			
Forgiveness	Low			
Compassion	Low			
Empathy	Low			
Kindness	Low			
Generosity	Low			
Humility	Low			
Patience	Low			
Self-control	Low			
Discipline	Low			
Responsibility	Low			
Accountability	Low			
Integrity	Low			
Honesty	Low			
Truthfulness	Low			
Transparency	Low			
Openness	Low			
Curiosity	Low			
Learning orientation	Low			
Growth mindset	Low			
Fixed mindset	High			
Resilience training	No			
Mindfulness practice	No			
Therapy sessions	No			
Support groups	No			
Peer support	No			
Online communities	No			
Mobile apps	No			
Wearable devices	No			
Smart home technology	No			
Internet usage	Low			
Digital literacy	Low			
Technology adoption	Low			
Innovation spirit	Low			
Risk-taking behavior	Low			
Entrepreneurial mindset	Low			
Leadership skills	Low			
Teamwork ability	Low			
Communication skills	Low			
Conflict resolution	Low			
Negotiation skills	Low			
Problem-solving skills	Low			
Decision-making skills	Low			
Time management	Low			
Organization skills	Low			
Planning abilities	Low			
Goal setting	Low			
Motivation levels	Low			
Energy levels	Low			
Vitality	Low			
Endurance	Low			
Strength	Low			
Flexibility	Low			
Balance	Low			
Coordination	Low			
Agility	Low			
Speed	Low			
Reaction time	Slow			
Reflexes	Slow			
Hand-eye coordination	Poor			
Fine motor skills	Impaired			
Gross motor skills	Impaired			
Posture	Poor			
Balance tests	Failed			
Coordination tests	Failed			
Agility tests	Failed			
Speed tests	Failed			
Reaction time tests	Failed			
Reflex tests	Failed			
Hand-eye coordination tests	Failed			
Fine motor skill tests	Failed			
Gross motor skill tests	Failed			
Posture assessments	Failed			
Balance assessments	Failed			
Coordination assessments	Failed			
Agility assessments	Failed			
Speed assessments	Failed			</

Db 801 WTTTWTAAAAAAATTTMMAGARAAKATWTWMAAAAAAAAMFRMAATATATTTT 860
 QY 612 ACCCTA-AAGATTACACCCAGATGTGATTAATCACTTCAAGAAAGACTTCCCA 670
 Db 861 ATATATTAATAAAAAATTAATAAARARRRKGAATAAAAAATTAATTTWTATTA 920
 QY 671 ATGAGCTCAACACTGCAGTTGTCTTCCCTGGCCAGTAGAGTACTGACAAATGCTG 730
 Db 921 AAAATTTTAMAMWMAAAAAAAMAAAAAGAAARAGAMAMATTTTATTAATRWMA 980
 QY 731 ATGATGGAGTGTGATGTGGAATGTCAAATACTAATACAGATCTATGCTCAAA 790
 Db 981 AAAAATTTTWTWAAAAAMWDTTATWTATWTATWATAGAMAAATADATWRRA 1040
 QY 791 CT-----AAACCACTACTGATTAATCAAAATCCTTCAACTTTAATTCAGAGCAATGC 844
 Db 1041 TTATVADAAARAAAGAAATKADAGATAAATTTTATTTTAAAAAAGAAAAA 1100
 QY 845 CTGTGTCACACATAGATACGATTCATGATGATGATGATGATGATGATGATGATG 904
 Db 1101 TTTTAAAAAATAAAAAAATRAAAAAAATAAAAAAAGAAKADAAVAGAAATTA 1160
 QY 905 TCCATTTAGATGAAAAA 921
 Db 1161 AAAAAAAGMRAARAW 1177

RESULT 3
 CNS05EVL 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
 DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069706 GI:4949849
 VERSION AL069706.1
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydriidae; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Direct Submission
 TITLE Genoscope.
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammosser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y?; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR29B23"
 /clone_lib="RPCI-98"
 /note="end : T7"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 2.1%; Score 66.4; DB 29; Length 1101;
 Best local Similarity 35.1%; Pred. No. 0.018;
 Matches 224; Conservative 104; Mismatches 300; Indels 10; Gaps 3;
 QY 4 AATATTCTGAAAAAAGCTTAAGATTAATACATGATAGTGATGATGATGATGATGATG 63
 Db 471 MAATWMAAAMAAATTAATWATWMAAAMAAAMAAAMAAATTTTMMWMTATTTWMAW 530
 QY 64 CTGTGCTTCGCAAGCTTTGGCTTTAAGCAATCGAATAGAGTAACGATTAACGCAATTA 123
 Db 531 TWTAMTAAAAAATAATTTTAAAMWMAATTAATTAATTAATTAATTAATTAATTA 590
 QY 124 GTTAATCAAGCAAGCAAGCGATGCTAATTCCTGTAAGCTGACGCTGTGCAAAAT 183
 Db 591 TATTAATWATTAATTAATTAATTAATTAATTTTWTATTAATTTTAATTAATTAAT 650
 QY 184 GGTTCGTTGTCATACAGCTTCTTAGAGATGATGATTAATTAACGACGCTAAT 243
 Db 651 TATTAATTAATWATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 710
 QY 244 GCACCAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 303
 Db 711 AATAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 770
 QY 304 GATTGTGTCGATACCAAGTAACAAATAGTTTCAGATTAACACACTAGCAGAAATAGA 363
 Db 771 TATWMAATWMAAMWMTATATWATTAATTAATTAATTAATTAATTAATTAATTA 828
 QY 364 TTTGATCAAAAGACAAACAGACATATTAATGCTGTGTTATGATGATGATGATGAT 423
 Db 829 TATWATWATATWATWATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 888
 QY 424 CATTTAAAAAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 483
 Db 889 TWTTTTMMWMAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 943
 QY 484 TTTGTAATGCTGTGCTGTTGATTAATCCAGCTCAGATTAATTAATTAATTAATTAAT 543
 Db 944 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1000
 QY 544 AATTTTAATTTTCAAAACCAACTCAAGCTGAATTTGTTATGATTAATTAATTAATTA 603
 Db 1001 AATATATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1060
 QY 604 CCAATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 641
 Db 1061 TTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1098

RESULT 4
 BX139987 844 bp DNA linear GSS 13-MAR-2003
 LOCUS Dario rerio genomic clone DKEY-99E7, genomic survey sequence.
 DEFINITION BX139987
 ACCESSION BX139987
 VERSION BX139987.1 GI:27971314
 KEYWORDS GSS.
 SOURCE Dario rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 844)
 HUMPHREY,S.J., HUCKLE,E. and DURHAM,J.L.
 Direct Submission
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk unpublished
 This sequence was generated from the T7 end of BAC 99E7. 99E7 is
 part of the Dariokey BAC library created by R. Plasterk and N.V.
 Keygene. Further details:
 http://www.sanger.ac.uk/Projects/D_rerio/
 location/Qualifiers

FEATURES

source
 1..844

```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-99E7"
/tissue_type="testis"
/notes="vector pindigBAC-536"

```

```

BASE COUNT      406 a      49 c      81 g      308 t
ORIGIN
Query Match      2.1%; Score 65.8; DB 29; Length 844;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 241; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

```

```

47 TACGCTATTGGAGCTCTGCTGCGACGCTTGGCTTTAGACATCAGATTAAGATA 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 TTCCAGCTCATTTATGCTTTTATCTTATGATTAATTAATTAATTAATTAATTA 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 ACGATTAACAGCAATAGTAAATCAAGCAAGACGCTAGATCTAATCTGTAGACCTG 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 CAGGCTTGGACAAATGGCTGCTGCTTAATACAGTCTTGAAGATGTTGATGACT 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 TTATTAACAGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTAATTA 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
492 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 551
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 GTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 611
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 CAATTAACAGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTAATTA 731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
407 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
672 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 GAATTAACAGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
732 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
792 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 5      1101 bp      DNA      linear      GSS 03-JUN-1999
CNS0039G      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      1101 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION      Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL063921
VERSION      AL063921.1 GI:4941778
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS      Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
COMMENT      Web : www.genoscope.cns.fr
The BACP is constructing a physical map of the Drosophila

```

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Kammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

```

1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_id="RPI-98"
/notes="end : TET3"
Location/Qualifiers
BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN
Query Match      2.0%; Score 65.2; DB 29; Length 1101;
Best Local Similarity 19.1%; Pred. No. 0.03;
Matches 129; Conservative 264; Mismatches 282; Indels 1; Gaps 1;

```

```

324 AAACAATAGTTTACATTAACACACTAGCAATAGATTGATCAAGCAACAG 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
419 AAACAATAGTTTACATTAACACACTAGCAATAGATTGATCAAGCAACAG 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 AGCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
564 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 623
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
624 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
718 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 743
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
778 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
744 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 803
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
838 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
898 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
864 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
958 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1017
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1077
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db	1078	MAMWHMHYAHYHHAH	1093
RESULT 6	BX173672	829 bp DNA	Linear GSS 13-MAR-2003
LOCUS	Danio rerio genomic clone DKEX-150M6,	genomic survey sequence.	
ACCESSION	BX173672		
VERSION	BX173672.1	GI:28005377	
KEYWORDS	GSS.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
REFERENCE	Humphray, S.J., Hucklee, E. and Durham, J.L.		
AUTHORS	Direct Submission		
TITLE	Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphraysanger.ac.uk Unpublished		
JOURNAL	This sequence was generated from the SP6 end of BAC 150M6. 150M6 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Koyne. Further details:		
COMMENT	http://www.sanger.ac.uk/Projects/D_rerio/ Location/Qualifiers		
FEATURES	source 1..829 /organism="Danio rerio" /mol_type="genomic DNA" /db_xref="taxon:7955" /clone="DKEX-150M6" /issue_type="Testis" /note="vector pindigobac-536"		
BASE COUNT	260 a 103 c 32 g 434 t		
ORIGIN			
Query Match	2.0%; Score 64.6; DB 29; Length 829;		
Best Local Similarity	45.5%; Pred. No. 0.038;		
Matches 229; Conservative 0; Mismatches 274; Indels 0; Gaps 0;			
OY	TTTAAACAATCAGATTAGAAGTAAGCAATGCGCATTTAGTATTCACGAAGAACCCTA	144	
DB	TTGAAACATTTTGAAAGAAAAACGTAAATATATATATTAAGATAATTAATATATGATGAT	610	
OY	GATGCTAATTCCTGTGACACTTGCCAGGCTTGGACAAAATGGTTCGTTCCAATACAGTT	204	
DB	GATGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	550	
OY	CTTAGAGATGTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	264	
DB	AATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	490	
OY	AGTTTCTAATACCATTAATATAGTTTATAGTATGATGATGATGATGATGATGATGATGAT	324	
DB	AGAAATATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	430	
OY	AAACAATAGTTTCAGATTACACACTAGCAAGATGATTTGCATCAAGACAACAGA	384	
DB	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	370	
OY	GCATATTAATGCTGTTGGTTAATGATGAAGCTAACGTTCAATTTAAAAAGATTAACT	444	
DB	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	310	
OY	AATCAATTAATTAATTTGGTAATAGAAACAATCTAAGTTTGAATTTGGTGGTGGAT	504	
DB	AATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	250	
OY	AATCACAGTCACGTAATTTAGATTACTGATGATGATGATGATGATGATGATGATGAT	564	
DB	GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	190	
OY	ACTCAAGTGAATTTGTTAATGA	587	

	Db	189	GATGATGATGATTAATAAATAAA	167
RESULT 7	CNS01VYG			
LOCUS	CNS01VYG	773 bp	DNA	linear
DEFINITION	Tetradodon nigroviridis genome survey sequence T7 end of clone 199C24 of library G from Tetradodon nigroviridis, genomic survey sequence.			
ACCESSION	AL169549			
VERSION	ALI69549.1	GI:7807606		
KEYWORDS	GSS: genome survey sequence.			
SOURCE	Tetradodon nigroviridis			
ORGANISM	Tetradodon nigroviridis			
REFERENCE	1			
AUTHORS	Roest Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fischer,C., Bernot,A., Fitzames,C., Mincker,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.			
TITLE	Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence			
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)			
MEDLINE	20296633			
PUBMED	10835645			
REFERENCE	2			
AUTHORS	Roest Crolius,H., Jalllon,O., Dasilva,C., Ozouf-Costaz,C., Fitzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.			
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis			
JOURNAL	Genome Res. 10 (7), 939-949 (2000)			
MEDLINE	20359837			
PUBMED	10899143			
REFERENCE	3	(bases 1 to 773)		
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
COMMENT	- Web : www.genoscope.cns.fr This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetradodon.			
FEATURES	Location/Qualifiers			
SOURCE	1..773			
	/organism="Tetradodon nigroviridis"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:99883"			
	/clone="199C24"			
	/clone_lib="G"			
BASE COUNT	/note="Genoscope sequence ID : COAG199BB12P1-end : T7"			
ORIGIN	210 a 265 c 69 g 184 t 45 others			
Query Match	2.0% ; Score 64.2; DB 29; Length 773;			
Best Local Similarity	43.9% ; Pred. No. 0.045;			
Matches 162; Conservative	22; Mismatches 185; Indels 0; Gaps 0;			
OY	1240 ACGTGATACGGTGACAACAACAAGCAACAACGACAGCTGATACATCATGAGTTCA 1299			
DB	165 ACTGCTACTACTACMACACTGCACACACACACACACACACACACACACACACAC 224			
OY	1300 ACAAGTCGCTCGAACAGAAATACACTACACACTTCTCAACAGATTCTTAATCCCTT 1359			
DB	225 AACTGCTGCTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 284			
OY	1360 AAATCTTATCTGCTATTGGAATGATGATGATGATGATGATGATGATGATGATG 1419			
DB	285 ACTGAC 344			

OY	1420	ACTAATTGGCGAGTCTCTAACGTTATGGACCAAGAATATATGCTGAATCCAGATTGGT	1479
Db	345	ACTACTCCTACTACTACTGCAMACCAGCATACATCTACTCTACTACTCATCTCTACAT	404
OY	1480	ATTCAAAATGGAATTCATATACTAATGACAGAACTTTATCGSAAACCAATTGGGT	1539
Db	405	ACTGCAACTACMGCAACTGCMACGTCRCAACMCWACTACTACCAWMACAACAATCWCW	464
OY	1540	GTTGGTTTACTTCAACAGGTTCAAGAGTAAGTTTAAGAGCTTCTTATAAGGGGAGCAA	1599
		: :	: :
Db	465	ACWBRAACTACTGCTCAACAATACTACTCMACTGCTACTACTACTACTGCTACTCTCT	524
OY	1600	CGTCCAAT	1608
Db	525	ACTACTACT	533

RESULT 8	1200 bp	mRNA	linear	EST 22-MAY-2003
EX447454	EX447454			
LOCUS	EX447454	Homo sapiens	THYMUS	Homo sapiens
DEFINITION	5-PRIME, mRNA sequence.			CDNA clone CS0CAP003Y119
ACCESSION	EX447454			
VERSION	EX447454.1	GI:31018746		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			

REFERENCE 1 (bases 1 to 1200)
 AUTHORS Li, W.B., Gruber, C., Jesssee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope

BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6207.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AAM22D03QPI&cluster=6207.r>. Contact : Feng Liang Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue sequence ID : CS0AAM22D03QPI.

FEATURES	Location/Qualifiers
source	1. 1200

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP003YH19"
/rissue_type="THYMUS"
/clone_idb="Homo sapiens THYMUS"
/vector="PCWVSPT_6; 1st strand cDNA was primed
with a NotI-Oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCWVSPT 6 vector.

```

```

Library was not normalized."
BASE COUNT      465 a      222 c      150 g      363 t
ORIGIN

```

Query Match	2.0%;	Score 63.8;	DB 13;	Length 1200;
Best Local Similarity	47.6%;	Pred. No. 0.057;		
Matches 188;	Conservative	0;	Mismatches 207;	Indels 0;
			Gaps	0

[illegible]

283 TATGCTTAGATCTAAGTGAATGCTGTGCGATACAAAGTAAACAATAGTTTCAGAT 342

Db 866 AAAAATTTCAAAATTTTAAATTTTATCAAAAAATAAAAAAATAAAAAATTCGATAT 922

QY	343	TACACACATAGCAGAAATAGATTGGATGATCAAGACAAACACAGACATTTAAGCTCTGTG	402
Db	926	GAATAAAATATATAATTAATTTGAAAAAATAAAATATATATTTTATGTAATTAA	985
QY	403	GTTATAGTGAAGCTAACGTTCAATTTAAAAAGAAATTAATACTAACCTCAATAGATTGGT	462
Db	966	TTTAAAAATAAAAAATTTAAAAAATTTAAAAATTTTGAACACATATAATTGGAA	1045
QY	463	AATGACAACAACATTTCTAAGTTGTAATGGGTGGTGGATATCCAGCTCAGCATATT	522
Db	1046	ATTAAAAATGATGATTAATAATGGATTAATTGAAAAATTTGAACACTGGATATTAAAAATT	1105
QY	523	AGATTACTAGTGATGGAGCTAAATTTAATTTTACACACAACTCAAGGTGAATTCCT	582
Db	1106	GAAATATATGATTAATTTGAAAAAATAAAATGAATAAAAAATAAAATTAATTTTTT	1165
QY	583	AATGACTCATTTTGTAGATGCCCACTCTACCTAA	617
Db	1166	AAATTTAAATTTAATTAACAAATATATTTTAAAAAA	1200

RESULT 9	FR0018463/c	470 bp	DNA	linear	GSS 09-DEC-1997
LOCUS	FR0018463				
DEFINITION	F.rubripes GSS sequence, clone 01E1Iac6,				genomic survey sequence.
ACCESSION	AL011359				
VERSION	AL011359.1				
KEYWORDS	GI:2676793				
	GSS: genome survey sequence.				

REFERENCE

1 (basses 1 to 470)

Urochordata

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neuteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neuteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.

AUTHOR: Williams, G. and Brenner, S.
 TITLE: Direct Submission
 JOURNAL: Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
 COMMENT: Centre Hinton, Cambridge, CB10 1SB. Email: d10nel@hmp.mrc.ac.uk
 Vector: pBluescript II KS
 v_type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES	Location/Qualifiers
source	1..470

Query Match	Score	DB	Length
1.9%	61.8	29	470
51.1%	Pred. No. 0.12		

QY	2920	AAAGTGAACCTGACGCTGCTGTGGTTCAGTTTACAGAGAATTATTACCAACT	2977
Bestc. local, 100%	Conservative	0	Mismatches 127; Indels 0; Gaps 0
Matches 135;			

[illegible]

Db
285 ACTACTGCTGCTACTGCTACTGCTGCTACTGCTACTGCTACTGCTGCTCTNCTAATNCT 226

[illegible]

OY	3100	GCTTAACCACTGGGCGCTTAATCTGGGCGGCTTCAAAACCAACTGCTCCTAACCAAGCT	3159
Db	165	GCTACTCTGCTGCTACTACTACTGCTGCTACTGCTACTACTGCTACTGCTGCTACTACT	106
OY	3160	GCTCCAAACCAACCAACCGCTCCCA	3181
Db	105	ACTGCTACTACTGCTGCTGCTA	84
RESULT 10			
B2643398		480 bp	DNA
LOCUS			linear
DEFINITION	OGAOW577C.ZM.0.7.1.5_KB zea mays genomic clone ZMBMa011J17,		GSS 29-JAN-2003
ACCESSION	B2643398		
VERSION	B2643398.1		
KEYWORDS	GSS.		
SOURCE	zea mays		
ORGANISM	zea mays		
REFERENCE	B2643398		
AUTHORS	OGAOW577C.ZM.0.7.1.5_KB zea mays genomic clone ZMBMa011J17,		
	genomic survey sequence.		
TITLE	Unpublished		
JOURNAL	Other GSSs: OGAOW577M		
COMMENT	Contact: Cathy Whiteleaw		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-838-5843		
	Fax: 301-838-0208		
	Email: whiteleaw@tigr.org		
	Seq primer: TF		
	Class: sheared ends.		
FEATURES			
source	Location/Qualifiers		
	1..480		
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
	/strain="B73"		
	/db_xref="taxon:4577"		
	/clone="ZMBMa011J17"		
	/clone_1kb="ZM.0.7.1.5_KB"		
	/note="Vector: pBSCSK-; Site 1: HincII; 0.7-1.5 kb		
	methylation filtered genomic DNA library"		
BASE COUNT	96 a 160 c 65 g 159 t		
ORIGIN			
Query Match	1.9%: Score 61.8; DB 29; Length 480;		
Best Local Similarity	52.5%: Pred. No. 0.12;		
Matches 135:	Conservative 0; Mismatches 122; Indels 0; Gaps 0;		
OY	2923	GTTGATACCTTGACTGCTGCTGTTGGTTCAGTTTACAGAAGATTATTACCAACTGCT	2982
Db	74	GCTACTACTACTACTGCTGCTGCTACTACTACTGCTAATGCTACTACTACTACTGCTGCT	133
OY	2983	AAGCTTAAGAAAAAAGCTGCTGCTTTTAGTGCTCTGGTAAATCTGGTGATAAAGAACTGCT	3042
Db	134	ACTACTACTACTACTGCTGCTGCTACAGTACTACTGCTACTGCTTCTACTACTACTACT	193
OY	3043	GCTGCTACTTAACCTGCTGCTGCTCAAGCTTAACCAATCTGGACCAAGAGTAGTCAACGCT	3102
Db	194	GCTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT	253
OY	3103	AAACCAACTGGGCTAAATCTGGTGGCGCTTACAAACCAACCACTGCTCTTAAGCAGAGCTCT	3162
Db	254	GCTACTACTGCTACTACTACTACTACTGCTACTGCTACTGCTACTACTACTACTACTGCTACT	313
OY	3163	CCAAACCAACCGCTCC	3179
Db	314	ACTGCTACTACTGCTACT	330

[illegible]

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	124.6	3.9	560073	4	US-08-545-5280-1	Sequence 1, App11	
2	78.4	2.5	4162	1	US-07-973-257-1	Sequence 1, App11	
3	60.2	1.9	3489	2	US-08-728-1234-1	Sequence 1, App11	
4	60.2	1.9	3489	4	US-09-298-568-1	Sequence 1, App11	
5	60.2	1.9	3489	2	US-09-410-199-1	Sequence 1, App11	
6	60.2	1.9	32207	2	US-08-770-379-20	Sequence 20, App1	
7	60.2	1.9	32207	3	US-08-757-6694-20	Sequence 20, App1	
8	60.2	1.9	32207	4	US-09-230-3714-20	Sequence 20, App1	
9	53.2	1.7	987	4	US-09-142-584-5	Sequence 5, App1	
10	53.2	1.7	1037	4	US-09-181-585-3	Sequence 3, App1	
11	53.2	1.7	1159	4	US-09-181-585-1	Sequence 1, App11	
12	53.2	1.7	1471	4	US-09-181-585-2	Sequence 2, App11	
13	53	1.6	987	4	US-09-142-584-3	Sequence 3, App11	
14	52.4	1.6	3057	4	US-09-601-198-55	Sequence 55, App1	
15	51.4	1.6	987	4	US-09-142-584-1	Sequence 1, App11	
16	51	1.6	636	3	US-08-998-416-1137	Sequence 1137, App1	
17	49.8	1.6	300	4	US-09-328-352-4065	Sequence 4065, App1	
18	49.4	1.5	1887	4	US-09-601-198-39	Sequence 39, App1	
19	49	1.5	5163	3	US-08-700-651-1	Sequence 1, App11	
20	49	1.5	5163	3	US-08-928-361B-4	Sequence 4, App11	
21	49	1.5	5163	4	US-09-588-995A-4	Sequence 4, App11	
22	49	1.5	5318	3	US-08-700-651-2	Sequence 2, App11	
23	49	1.5	5318	3	US-08-928-361B-3	Sequence 3, App11	
24	49	1.5	5318	4	US-09-588-995A-3	Sequence 3, App11	
25	48.4	1.5	397	3	US-09-253-691-3	Sequence 3, App1	
26	48.4	1.5	4864	3	US-07-665-792E-10	Sequence 10, App1	
27	48.2	1.5	234	1	US-08-469-802B-3	Sequence 3, App11	

C	28	48.2	1.5	224	2	US-08-267-803B-3	Sequence 3, Appl
C	29	47.6	1.5	195	1	US-08-469-802B-2	Sequence 2, Appl
C	30	47.6	1.5	195	2	US-08-267-803B-1	Sequence 2, Appl
C	31	47.4	1.5	1664976	4	US-08-916-441B-1	Sequence 1, Appl
C	32	47.2	1.5	165	3	US-09-043-303-17	Sequence 17, Appl
C	33	47.2	1.5	203	3	US-09-043-303-7	Sequence 7, Appl
C	34	46.2	1.4	5511	3	US-08-928-361B-2	Sequence 2, Appl
C	35	46.2	1.4	5511	4	US-09-588-995A-2	Sequence 2, Appl
C	36	46.2	1.4	7334	4	US-08-928-361B-1	Sequence 1, Appl
C	37	46.2	1.4	7334	4	US-09-588-995A-1	Sequence 10, Appl
C	38	46	1.4	5181	4	US-08-257-073-10	Sequence 208, App
C	39	46	1.4	16442	4	US-08-781-891-208	Sequence 208, App
C	40	46	1.4	16442	4	US-08-618-166-208	Sequence 27, Appl
C	41	45.8	1.4	4818	3	US-08-817-926-27	Sequence 1, Appl
C	42	45.4	1.4	2576	3	US-08-669-408B-1	Sequence 56, Appl
C	43	45.4	1.4	14066	4	US-09-601-198-56	Sequence 8, Appl
C	44	45.2	1.4	1555	3	US-08-669-408B-9	Sequence 21, Appl
C	45	45	1.4	5340	4	US-09-627-122-21	

ALIGNMENTS

```

RESULT 1
US-08-545-528D-1
: Sequence 1, Application US/08545528D
: Patent No. 6537773
: GENERAL INFORMATION:
: APPLICANT: Fraser et al.
: TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Frag
: Patent No. 6537773
: TITLE OF INVENTION: Thereof, and Uses Thereof
: FILE REFERENCE: P4193P1
: CURRENT APPLICATION NUMBER: US/08/545,528D
: CURRENT FILING DATE: 1995-10-19
: PRIOR APPLICATION NUMBER: US 08/488,018
: PRIOR FILING DATE: 1995-06-07
: PRIOR APPLICATION NUMBER: US 08/473,545
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 580073
: TYPE: DNA
: ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match      3.9%; Score 124.6; DB 4; Length 580073;
Best Local Similarity 63.5%; Pred. No. 1.9e-18;
Matches 190; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 2770 TTTGGCCCTTACTTCATGGGTAGTGCCTCAGACGAATTTGGTAGACATTAGTATTTCT 2829
Db 228703 TTTTCACGACACCCCTGCATGGGTGATGCCCTGTATAGTAGGTTCTTCAGTTGGATCTTG 228762

QY 2830 GCAATTTATGATCATCTTATAGATTAGCTATTCGTTATTCCTTTAAGAGCTCAAGAAATTA 2889
Db 228763 TTTATCTTGTAGTCTTATGACTTTAGACTTTGGATTTGGATTCACCATTTACAGGGTAGAAACATC 228822

QY 2890 CAAGACAAAGGGCTTCAAAACAACATTTCAAAAAAGTTGATTTACCTTGACTGCTGCTTGGT 2949
Db 228823 CAAGATGCAATCGTTGTGTTAAAGTCTTTTAAAAAGTTGATACACTACAACATCTGTCGGT 228882

QY 2950 TCAGTTTACAGAAGATATTATTAACCAACTGCTAAGCTTTAAGAAAAAAACCTGCTGTTTA 3009
Db 228883 AGTGTGACAAAAAGATTATTAACCAACTGCTGTGGTGAAGAAAAAGACACACTGTGCATTG 228942

QY 3010 GGTGCTGTAATCTGTGTATAGAAACCTGCTGCTGCTGCTAATCACTGCTGCTCCAGC 3068
Db 228943 AAAGCTGTAATCTGACTAGTGTAAAGAAAAAACCCTGCTCTTTTAAAAACACCTGTTCAAC 229001

RESULT 2

```

```

US-07-973-257-1
; Sequence 1, Application US/07973257
; Patent No. 5378820
; GENERAL INFORMATION:
; APPLICANT: Keeler, Jr. Calvin L.
; APPLICANT: Dohms, John E.
; TITLE OF INVENTION: Gene Encoding Cytohesin
; TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly and Hutz
; STREET: 1220 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM/PC or Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/973,257
; FILING DATE: 19921109
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: No. 5378820e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4182 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma Gallisepticum
; STRAIN: S6
US-07-973-257-1

Query Match          2.5%; Score 78.4; DB 1; Length 4182;
Best Local Similarity 98.8%; Pred. No. 1.2e-08;
Matches 79; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATGAATATTTCTAAACCTTAAGTTATACATGATGATGATGATGATGATTTGGA 60
      |||
Db      4103 ATGAATATTTCTAAACCTTAAGTTATACATGATGATGATGATGATTTGGA 4162

QY      61 GCTCTTGTTCTGCAAGCTT 80
      |||
Db      4163 ACTCTTGTTCTGCAAGCTT 4182

RESULT 3
US-08-728-323A-1/C
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```

```

; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPMW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match          1.9%; Score 60.2; DB 2; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.00017;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY      2998 CCTGCTGCTTATAGTGGTGAATCTGATGATAGTAAGAACCTGCTGCTGAACCT 3057
      |||
Db      2236 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2177

QY      3058 GCTGCTCAGCTAAACCTATGCTGACCAACGTAAGCTACCACTAAACCACTGGGC-- 3115
      |||
Db      2176 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117

QY      3116 -CTAATCTGGGCGCCCTGACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCT 3174
      |||
Db      2116 GCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2057

QY      3175 GCTCC 3179
      |||
Db      2056 GCTCC 2052

RESULT 4
US-09-298-568-1/C
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

```


Db 1099 CTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1158
QY 3050 CTAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3109
Db 1159 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1218
QY 3110 CTGGGCTTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3169
Db 1219 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278
QY 3170 CAACGCTCC 3179
Db 1279 CTGCTGCTGC 1288

RESULT 13
US-09-142-584-3
; Sequence 3, Application US/09142584
; Patent No. 6403094
; GENERAL INFORMATION:
; APPLICANT: Tilbail, Richard W.
; APPLICANT: Williamson, Ethel D.
; APPLICANT: Hayward, Helen L.
; APPLICANT: Oyston, Petra C.F.
; APPLICANT: Payne, Dean W.
; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES
; FILE REFERENCE: 124-665
; CURRENT APPLICATION NUMBER: US/09/142,584
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/GB97/00660
; EARLIER FILING DATE: 1997-03-11
; EARLIER APPLICATION NUMBER: GB 9605222.0
; EARLIER FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: (136)..(987)
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: (1)..(32)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(984)
; US-09-142-584-3

Query Match 1.7% Score 53; DB 4; Length 987;
Best Local Similarity 44.6%; Pred. No. 0.0048;
Matches 209; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 363 ATTGTGCAAGCAAGCAAGCAAGCAATATATGCTCTGTTGTTATGATGAGCAATACGT 422
Db 117 ATCTAATGAATGTCACAAAGCAAGCTCTTATGATATGATGATCACTTAATTTAGAAAG 176
QY 423 TCATTTAAAGAAATTAATCTACTACATCAATAGATTTGTAATAGAAACCAATTTCTAA 482
Db 177 AAGATTAATTAACAAATATATTAATTAAGAGAAATGAAATATTTATCTTAATGCTAT 236
QY 483 GTTGTAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 542
Db 237 GGCATATTTGATAGTTACTATTAATTCACAAAGAAATGATTTTATATTAATATATTC 296
QY 543 TAAATTTAATTTTACAAACCAACTCAAGTGAATTTGTTAAAGACTTCAATTTTAGATGC 602
Db 297 TAAAGTTGAATTAATGAGAACCAATCAATGAATTAATCTGAAGATCTTTATGTTGAAA 356
QY 603 GCCAATCTTAACCAATTAATTAACCAAGATTTGTTAATCAATTTAATCAATCAATTCAC 662
Db 357 AGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 416

QY 663 CTACCAATGACCTCAACACTGACATTTGTTCTTGCCAGTAGTAGATGATGAGAAC 722
Db 417 TAAATATACATGATACAGTACATGACATGACATGACATGACATGACATGACATGAC 476
QY 723 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 782
Db 477 AACTGCTAATGTTACTGTTCTTTTAAATGAACAGAGATTAATTAATCACTAGTTATAG 536
QY 783 TGCTCAACTAATCAACACTACTGATTAATCAAAATCTTCAACTTTTAAAT 831
Db 537 TTTGCAAAATACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 585

RESULT 14
US-09-601-198-55
; Sequence 55, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Helner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 3057
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
; US-09-601-198-55

Query Match 1.6% Score 52.4; DB 4; Length 3057;
Best Local Similarity 45.6%; Pred. No. 0.01;
Matches 220; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

QY 754 AATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 813
Db 1175 AATCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1234
QY 814 AATCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 873
Db 1235 TATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1294
QY 874 TGAATGTCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 933
Db 1295 CTACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1354
QY 934 GATGACTGTTCTGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 993
Db 1355 AATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1414
QY 994 AGCAACGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1053
Db 1415 GCATCTTTGCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1474
QY 1054 GATTCAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1113
Db 1475 ATTGCTCAATTTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
QY 1114 GGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1172
Db 1535 GGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1594
QY 1173 CCACATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1232

Db 1595 CTTATTCTTTTATGTAATATCCATCTCTTCATCAAGATTACTAATCATCTTTATA 1654
QY 1233 TT 1234
Db 1655 TT 1656

RESULT 15

US-09-142-584-1
; Sequence 1, Application US/09142584

; Patent No. 6403094

; GENERAL INFORMATION:

; APPLICANT: Tilball, Richard W.

; APPLICANT: Williamson, Ethel D.

; APPLICANT: Hayward, Helen L.

; APPLICANT: Oyston, Petra C.F.

; APPLICANT: Payne, Dean W.

; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES

; FILE REFERENCE: 124-665

; CURRENT APPLICATION NUMBER: US/09/142,584

; EARLIER APPLICATION NUMBER: PCT/GB97/00660

; EARLIER FILING DATE: 1997-03-11

; EARLIER APPLICATION NUMBER: GB 9605222.0

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 1

; LENGTH: 987

; TYPE: DNA

; ORGANISM: Clostridium perfringens

; FEATURE:

; NAME/KEY: mat_peptide

; LOCATION: (136)..(456)

; FEATURE:

; NAME/KEY: misc_signal

; LOCATION: (1)..(32)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(984).

; US-09-142-584-1

Query Match

Best Local Similarity 44.38; Pred. No. 0.011; Length 987;

Matches 208; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 363 ATTTGATCAAGACAAAGACATATTAATGCTGCTGTTATGATGAAGCTAAGT 422
Db 117 ATCTAATGAATGTCMAAAGCTCTCTTAATGATTAATGATTAATGATTAATG 176
QY 423 TCATTAAAGAAATTAATCTAATCAATAGATGTAATAGAAACAAATTTCTAA 482
Db 177 AAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 236
QY 483 GTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
Db 237 GGCATATTTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 296
QY 543 TAAATTTAATTTACAAACCAACCAAGTGAATGTAATGATTAATGATTAATG 602
Db 297 TAAAGTGAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 356
QY 603 GCCAATCTTACCTTAAGATTTACACCCAGATGTAATGATTAATGATTAATG 662
Db 357 AGCTCTCTTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 416
QY 663 CTTACCAATAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 722
Db 417 TAAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 476
QY 723 AAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 782

Db 477 AACTGCTAAGTTACTGTTCTTTTAATGAACAGAGATCAATTAACCTAGTTATAG 536
QY 783 TGCTCAACTAAACACCTACTGATTAATCAAAATCCCTCACTTTAAT 831
Db 537 TTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 585

Search completed: October 10, 2003, 11:49:06
Job time : 183 secs

THIS PAGE BLANK (USPTO)